NICE National Institute for Health and Care Excellence

Single Technology Appraisal

Empagliflozin for treating chronic heart failure with reduced ejection fraction [ID3826]

Committee Papers

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

SINGLE TECHNOLOGY APPRAISAL

Empagliflozin for treating chronic heart failure with reduced ejection fraction [ID3826]

Contents:

The following documents are made available to consultees and commentators:

The **final scope** and **final stakeholder list** are available on the NICE website.

1. **Company submission** from Boehringer Ingelheim
2. **Clarification questions and company responses**
3. **Patient group, professional group and NHS organisation submissions** from:
   a. Cardiomyopathy UK
   b. The British Society for Heart Failure
4. **Evidence Review Group report** prepared by BMJ Group
5. **Evidence Review Group report – factual accuracy check**
6. **Technical engagement response** from Boehringer Ingelheim
7. **Technical engagement responses and statements from experts:**
   a. Nick Hartshorne-Evans – patient expert, nominated by Pumping Marvellous Foundation
8. **Technical engagement responses from consultees and commentators:**
   a. AstraZeneca
9. **Evidence Review Group critique of company response to technical engagement** prepared by BMJ Group

Any information supplied to NICE which has been marked as confidential, has been redacted. All personal information has also been redacted.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Empagliflozin for treating chronic heart failure with reduced ejection fraction [ID3826]

Document B

Company evidence submission

29th June 2021

File name	Version	Contains confidential information	Date
ID3826_Empagliflozin_Document B_v7.0_AIC_CIC.docx	**7.0**	**Yes**	**29th June 2021**

Contents

Tables and figures

List of abbreviations

Acronym	Definition
ACCF	American college of cardiology foundation
ACE	Angiotensin-converting enzyme
ACEi	Angiotensin-converting enzyme inhibitor
AE	Adverse event
AESI	Adverse event of special interest
Afib	Atrial fibrillation
AHA	American heart association
AICc	Akaike Information Criterion with a correction for finite sample size
ARB	Angiotensin receptor blockers
ARNI	Angiotensin receptor-neprilysin inhibitor
BB	Beta-blocker
BIC	Bayesian Information Criterion
BNP	B-type natriuretic peptide
CEA	Cost-effectiveness analysis
CI	Confidence interval
CKD	Chronic kidney disease
CKD-EPI	Chronic kidney disease - Epidemiology collaboration equation
CMR	Cardiac magnetic resonance
COPD	Chronic obstructive pulmonary disease
CPET	Cardiopulmonary exercise test
CRM	Cardio-renal-metabolic
CRS	Cardiorenal syndrome
CSR	Clinical study report
CT	Computerized tomography
CV	Cardiovascular
DM	Diabetes mellitus
ECG	Electrocardiogram
eGFR	Estimated glomerular filtration rate
EF	Ejection fraction
ESC	European Society of Cardiology
EMA	European medicines agency
EQ-5D	EuroQol- 5 dimension
FCE	Finished Consultant Episodes
GEE	Generalised estimating equations
HbA1c	Glycated haemoglobin
HFmEF	Heart failure with mid-range ejection fraction
HFpEF	Heart failure with preserved ejection fraction
HFrEF	Heart failure with reduced ejection fraction

Acronym	Definition
HHF	Hospitalisation for heart failure
HR	Hazard ratio
HRG	Healthcare Resource Group
HRQoL	Health-related quality of life
IHD	Ischaemic heart disease
ITC	Indirect treatment comparison
KCCQ	Kansas City Cardiomyopathy Questionnaire
KCCQ-CSS	Kansas City Cardiomyopathy Questionnaire clinical summary score
KCCQ-TSS	Kansas City Cardiomyopathy Questionnaire total symptom score
KCCQ-OSS	Kansas City Cardiomyopathy Questionnaire overall summary score
LLA	Lower limb amputation
LVEF	Left ventricular ejection fraction
MedDRA	Medical dictionary for regulatory activities
MI	Myocardial infarction
MMRM	Mixed model for repeated measures
MOA	Mechanism of action
MRA	Mineralocorticoid receptor antagonists
MRI	Magnetic resonance imaging
MTD	Maximum tolerated dose
NICE	National institute for health and care excellence
NT-pro-BNP	N-terminal pro hormone B-type natriuretic peptide
NYHA	New York heart association
PT	Preferred term
QoL	Quality of life
RCT	Randomised controlled trial
RS	Randomised set
RWE	Real world evidence
SAE	Serious adverse event
SGLT1	Sodium-glucose co-transporter-1
SGLT2	Sodium-glucose co-transporter-2
SGLT2i	Sodium-glucose co-transporter-2 inhibitor
SmPC	Summary of medicinal product characteristics
SoC	Standard of care
SoC	System organ class
T2DM	Type 2 diabetes mellitus
TS	Treated set
UK	United Kingdom

B.1 Decision problem, description of the technology and clinical care pathway

B.1.1 Decision problem

The population defined in the Final Scope is an extension to the full marketing authorisation of empagliflozin that was issued by the European Medicines Agency (EMA) on 22 May 2014 (1). Currently, empagliflozin is indicated for the treatment of adults with insufficiently controlled type 2 diabetes mellitus (T2DM) as an adjunct to diet and exercise, to be used as a monotherapy when metformin is considered inappropriate due to intolerance or in addition to other medicinal products for the treatment of diabetes (1). Empagliflozin has also received a NICE recommendation for this indication (2, 3).

This submission covers the extension to the empagliflozin indication to include the treatment of adults with symptomatic heart failure (HF) and reduced ejection fraction (HFrEF). The cost-effectiveness, comparative effectiveness, clinical efficacy, and safety of empagliflozin *versus* standard care in adult patients with HFrEF are topics which are specifically addressed in this submission.

The company submission differs from the final NICE scope with regards to the comparators in the cost-effectiveness analysis, but is consistent with the evidence considered for decision making in the dapagliflozin appraisal ID1656 published on 24 December 2020 (4), as outlined in Table 1.

Table 1. The decision problem

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope
Population	Adults for the treatment of symptomatic chronic heart failure with reduced ejection fraction	Same	Not applicable
Intervention	Empagliflozin in combination with standard care (including diuretics, treatment with an ACE inhibitor, ARBs, mineralocorticoid receptor antagonist, beta-blockers, cardiac devices and sacubitril valsartan)	Same	Not applicable
Comparator(s)	• Individually optimised standard care without empagliflozin. Standard care is defined as: o ACE inhibitors in combination with beta-blockers, and/or mineralocorticoid receptor antagonists o ARBs in combination with beta-blockers, and/or mineralocorticoid receptor antagonists o Sacubitril valsartan in combination with beta-blockers, and/or mineralocorticoid receptor antagonists • Dapagliflozin as an add-on to standard care	Same, however dapagliflozin does not reflect current standard of care and is not a relevant comparator. The evidence for empagliflozin vs standard treatment with ACEi, ARBs, sacubitril/valsartan is the most relevant for the committee to consider. This is because a majority of eligible patients in the UK receive at least one of these products. Comparative analyses of empagliflozin vs dapagliflozin are provided in B.2.8 upon the request of the ERG and NICE Technical team; however these are secondary. This is consistent	The estimated prescribing of dapagliflozin is [redacted] in MQT May 2021 for patients with HF only. In the HF only population, it is prescribed [redacted] times less often than sacubitril valsartan (Table 2), a product considered as SoC but prescribed less frequently than ACEi and ARBs(5). The NICE dapagliflozin resource impact template estimated that 75% of HFrEF patients optimised on standard care with either ACE/ARBs, or sacubitril/valsartan and with an eGFR >30mL/min per 1.73m^2 will receive dapagliflozin by 2025. In 2021, 2022, 2023, 2024, the uptake is estimated to be 20%, 35%, 50%, and 60%, respectively (6). The resource impact template only considers patients with HFrEF only and not those with comorbid T2DM. There is limited empirical evidence to support these estimates. Given the market share in MQT May 2021 was [redacted] (5), it's unlikely that 20% of eligible HFrEF only patients would receive dapagliflozin by the end of 2021.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope
		with the perspective of UK clinicians we have consulted.	Future use of treatment is speculative. We can only reflect the care pathway used today in the submission. This is consistent with NICE guidance and committee discussion in TA398 (Section 4.18) where a specific future scenario was proposed by the manufacturer but rejected by the committee (7). It is clear from the data presented above that prescribing of dapagliflozin is not SoC at the time of submission, there is minimal scope for displacement, and hence we do not regard this comparator of primary relevance to the decision problem. Direct economic evidence for empagliflozin vs dapagliflozin is not informative. More important is patient and prescriber choice. As the key clinical efficacy outcomes for empagliflozin and dapagliflozin are comparable, the cost-effectiveness of SGLT2i vs SoC and is the most relevant economic evidence to consider, consistent with (8) 5.1.14 of the NICE Guide to Methods 2013. This is described in B.3.8.3.
Outcomes	The outcome measures to be considered include: • symptoms of heart failure • hospitalisation for heart failure • all-cause hospitalisation • mortality • cardiovascular mortality • kidney function • adverse effects of treatment	Same	Not applicable

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope
	• health-related quality of life		
Economic analysis	The reference case stipulates that the cost-effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. If the technology is likely to provide similar or greater health benefits at similar or lower cost than technologies recommended in published NICE technology appraisal guidance for the same indication, a cost-comparison may be carried out. The reference case stipulates that the time horizon for estimating clinical and cost-effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective. The availability of any commercial arrangements for the intervention, comparator and subsequent treatment technologies will be considered.	Same	Not applicable

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope
	The cost of background therapies, such as diuretics for people with oedema, should also be included in cost-effectiveness analyses.		
Subgroups to be considered	Not included in the draft scope	No subgroups were considered separately in the economic analysis	Not applicable
Special considerations including issues related to equity or equality	Not included in the draft scope	Broad prescribing of SGLT2is in primary and secondary care could reduce the inequality in access to heart failure care in the UK	The socio-economic inequalities in CV disease present a major and persistent UK public health challenge. The UK-based population studies demonstrate that socio-economic deprivation is a strong risk factor for the development of HF and adverse HF outcomes (9, 10). Individuals in the lowest socio-economic group are 1.61 times more likely to experience incident HF than the most affluent individuals and do so, on average, at a 3.5 years younger age with a greater comorbidity burden at time of HF symptom onset (9).

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope
			Socio-economic status has an impact on access to secondary care in the UK, and subsequently access to HF treatments. Moscelli et al. reported a statistically significant difference in waiting times across socio-economic groups for patients who attend the same hospital: patients living in more income deprived areas waited longer (35% difference, or 43 days) than patients who lived in less deprived areas. As well as waiting longer, coronary heart disease patients in a lower socio-economic class were admitted to hospital less often than those in a higher class (11). McCartney et al. reported on a prospective study of 7049 men and 8353 women in the west of Scotland followed up for 37 years. The likelihood of a hospital admission for cardiovascular disease was 21% higher for female patients in highest socio-economic class than patients in lowest class. Those patients in class IV and V also stayed 25% longer in hospital (589 vs. 736 bed day/1000 person years, respectively) (12). These studies indicate that if patients in lower socio-economic classes utilise secondary care less often, their opportunity to access HF medications would also be lower if they are solely prescribed in secondary care.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope
			BI support the UK Government's and NICE's commitment to the reduction of health inequalities, reiterated in the recent NICE five-year strategy publication (13, 14). Principle 9 of NICE's Social Value Judgments states that due regard must be given to reducing inequalities. It states that equality should be considered in relation to the 9 protected characteristics in the Equality Act 2010 (age, disability, gender reassignment, race, religion or belief, sex, sexual orientation, marriage and civil partnership, pregnancy and maternity) and socio-demographic factors (14). Further the COVID-19 Marmot review aims to reduce the widened gap in health inequalities and build a fairer society post pandemic.(15) Broad prescribing of SGLT2i across primary and secondary care can support the reduction in disparity in access to HF care across socio-economic groups within the UK. Only permitting a cardiologist to initiate a SGLT2i would likely widen the gap in health inequalities and lead to a delay in prescribing due to the limited resource in secondary care.

Table 2. Relative market share of dapagliflozin vs SoC for HF (May 2021) (5)

IQVIA LPD data	Patient population				
		June 2020 to Aug 2020	Sep to Nov 2020	Dec 2020 to Feb 2021	Mar to May 2021
England population	T2DM+HF	[redacted]	[redacted]	[redacted]	[redacted]
	HF only	[redacted]	[redacted]	[redacted]	[redacted]
	TOTAL	[redacted]	[redacted]	[redacted]	[redacted]
Market share					
ACE (%)	T2DM+HF	[redacted]	[redacted]	[redacted]	[redacted]
	HF only	[redacted]	[redacted]	[redacted]	[redacted]
ARB (%)	T2DM+HF	[redacted]	[redacted]	[redacted]	[redacted]
	HF only	[redacted]	[redacted]	[redacted]	[redacted]
MRAs	T2DM+HF	[redacted]	[redacted]	[redacted]	[redacted]
	HF only	[redacted]	[redacted]	[redacted]	[redacted]
Beta-blocker (%)	T2DM+HF	[redacted]	[redacted]	[redacted]	[redacted]
	HF only	[redacted]	[redacted]	[redacted]	[redacted]
Sacubitril valsartan, n (%)	T2DM+HF	[redacted]	[redacted]	[redacted]	[redacted]
	HF only	[redacted]	[redacted]	[redacted]	[redacted]
Dapagliflozin, n (%)	T2DM+HF	[redacted]	[redacted]	[redacted]	[redacted]
	HF only	[redacted]	[redacted]	[redacted]	[redacted]
Empagliflozin, n (%)	T2DM+HF	[redacted]	[redacted]	[redacted]	[redacted]
	HF only	[redacted]	[redacted]	[redacted]	[redacted]

a. The IQVIA LPD data estimates the percentage market share for all HF patient receiving an individual treatment and makes no distinction on whether they are symptomized, optimised or type (HFrEF, HFpEF or HFmEF). The study samples from 150 eligible GP practices and extrapolates nationally (UK). The estimates presented here are for the England population which represents 84.4% of the total UK population.
b. The definition of HF used for the IQVIA LPD study is broad and includes HF of any type with and without T2DM. The QOF code used for HF were G58 [heart failure], G5y4, G1yz1 [rheumatic left ventricular failure], 662f [NYHA I], 662g [NYHA II], 662h (NYHA III], 662i [NYHA IV]
c. A narrow definition of HF, consistent with the dapagliflozin marketing authorisation, might lead to a biased sample as QOF codes are sometimes missing or patients are miss-classified
d. The IQVIA Hospital Pharmacy Audit (HPA), a secondary care database, indicates that prescribing of dapagliflozin is predominantly in the community. Of the packs of dapagliflozin sold in Feb 2021, [redacted] % were dispensed in a community compared to [redacted] % in a hospital pharmacy. This indicates that the IQVIA LPD data is a representative sample and reflects our best available evidence on the market share of dapagliflozin.

B.1.2 Description of the technology being appraised

- Empagliflozin is an orally bioavailable, selective sodium-glucose co-transporter-2 inhibitor (SGLT2i) which has cardioprotective effects and improves heart failure-related outcomes (16, 17).

Empagliflozin's mechanism of action, marketing authorisation, indication, mode of administration and list price are summarised in Table 3. Appendix C includes the draft summary of product characteristics (SmPC) for empagliflozin.

Table 3: Technology being appraised

UK approved name and brand name	Empagliflozin (Jardiance®)
Mechanism of action	Empagliflozin is an orally bioavailable, reversible, highly potent and selective inhibitor of SGLT2 (16). Through SGLT2 inhibition, empagliflozin simultaneously reduces renal reabsorption of glucose and sodium in the proximal tubules of the kidney and leads to increased urinary excretion of glucose and moderate natriuresis. The molecular bases of empagliflozin's cardioprotective and nephroprotective effects are unknown, but accumulating evidence suggests several distinct mechanisms are involved, including: • osmotic diuresis and natriuresis resulting in lowering of arterial pressure and stiffness and improvement in ventricular loading • improved myocardial and renal metabolism via switch to ketone bodies as the energy source • prevention of adverse cardiac remodelling through inhibition of inflammation, fibrosis, and cardiomyocyte cell death • direct inhibition of the Na+/H+ exchanger in myocardium, leading to reduction or reversal of cardiac injury, fibrosis, and systolic dysfunction • prevention of ischemia/reperfusion injury through decrease in calmodulin kinase II activity (17, 18).
Marketing authorisation/ CE mark status	Empagliflozin currently holds the EMA marketing authorisation and is recommended by NICE for the treatment of type 2 diabetes mellitus as a monotherapy (25 May 2016) or as a combination therapy with insulin or other antidiabetic drugs (25 March 2015) (1-3). On 20 May 2021, the Committee for Medicinal Products for Human Use (CHMP) adopted a positive opinion recommending a change to the terms of the marketing authorisation for the medicinal product Jardiance.

	A submission was also made to the MHRA, via the reliance route, on [redacted]. A UK MHRA Marketing Authorisation for HFrEF is expected w/c [redacted]. The draft SmPC is provided in Appendix C.
Indications and any restriction(s) as described in the summary of product characteristics (SmPC)	***Indication relevant to this submission***: Empagliflozin is indicated in adults for the treatment of symptomatic chronic heart failure with reduced ejection fraction. ***Other indications:*** Empagliflozin is indicated for the treatment of adults with insufficiently controlled type 2 diabetes mellitus as an adjunct to diet and exercise • as monotherapy when metformin is considered inappropriate due to intolerance • in addition to other medicinal products for the treatment of diabetes (19)
Method of administration and dosage	10 mg oral empagliflozin once daily
Additional tests or investigations	None
List price and average cost of a course of treatment	List price of a pack of 28 tablets (10mg) is £36.59. This equates to a cost of £1.31 per tablet per day for each patient.
Patient access scheme (if applicable)	None

Abbreviations: CHMP, Committee for Medicinal Products for Human use; EF, ejection fraction; EMA, European Medicines Agency; HF, heart failure; SGLT2i, sodium-glucose co-transporter 2 inhibitors.

B.1.3 Health condition and position of the technology in the treatment pathway

B.1.3.1 Overview of the disease for which the technology is indicated

- Heart failure is a complex clinical syndrome caused by structural and/or functional abnormalities of the myocardium resulting in the impairment of ventricular filling and ejection of blood (20, 21).
- Heart failure is classified based on left ventricular ejection fraction as HF with reduced, preserved or mid-range ejection fraction (HFrEF, HFpEF and HFmEF) (20, 22).
- Clinicians categorise heart failure according to the severity of the disease, structural changes, and symptoms using the two most common classification systems (American College of Cardiology Foundation (ACCF)/American Heart Association (AHA) and New York Heart Association (NYHA)) (20, 22, 23).

B.1.3.1.1 Disease overview

Clinical presentation and aetiology of heart failure

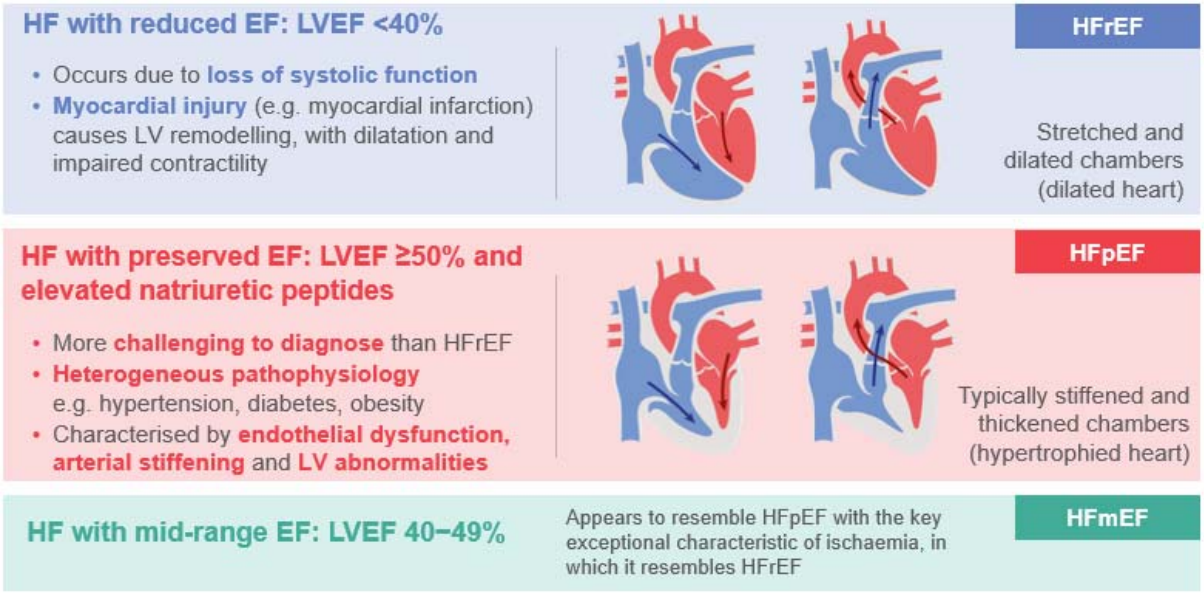
Heart failure (HF) is a complex clinical syndrome caused by structural and/or functional cardiac abnormality that results in reduced cardiac output and/or elevated intracardiac pressure, impairing the ability of the heart to function adequately and act as a pump to support physiological circulation (20, 21). Heart failure is characterised by a range of symptoms including breathlessness, fatigue, poor exercise tolerance, ankle swelling and peripheral oedema, however none are specific to HF (21). Signs of congestion, such as jugular venous distention, gallop rhythm and displaced apical impulse, are more specific to HF and indicative of higher risk of adverse outcome although harder to detect (20, 21). Heart failure results from injury to the myocardium caused by a wide range of pathologies including ischaemic heart disease, congenital heart defects, hypertension, and non-cardiovascular systemic diseases such as diabetes and severe lung disease (24). More than two-thirds of all cases of HF can be attributed to ischaemic heart disease, hypertension (25), obesity, chronic obstructive pulmonary disease (COPD) and rheumatic heart disease (24, 25). Less common aetiologies include cardiomyopathies, valvular disease, myocarditis, infections, systemic toxins, and cardiotoxic drugs (24).

Classification

Heart failure can be classified into acute and chronic in nature (20, 26). Acute HF (not relevant to the decision problem) is a life threatening condition, with a rapid onset of HF symptoms, typically leading to urgent hospital admissions (20, 27). Chronic HF (relevant to the decision problem) refers to patients who have had HF diagnosis for at least three months and can be categorised into left or right ventricular failure (28). Most patients have a systolic dysfunction caused by the left ventricle failing to pump blood efficiently. Failure of the left ventricle can also lead to right ventricular dysfunction by multiple mechanisms including myocardial ischemia involving both ventricles, increased pulmonary venous and arterial pressure, and reduced right ventricular coronary perfusion due to decreased systolic blood pressure (29). Based on the measurement of the left ventricular ejection fraction (LVEF), HF can be subcategorised into HF with preserved (HFpEF), mid-range (HFmrEF) and reduced ejection fraction

(HFrEF) as described in Figure 1. The focus of this submission is on the population with symptomatic HFrEF (20).

Figure 1. Left ventricular chronic heart failure



Abbreviations: EF, ejection fraction; HF, heart failure; HFmEF, heart failure with mid-range ejection fraction; HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; LV, left ventricular; LVEF, left ventricular ejection fraction.

Reference: (20, 22, 28, 30, 31)

Furthermore, clinicians often classify HF based on structural changes and symptom severity by using either the American College of Cardiology Foundation (ACCF)/American Heart Association (AHA) stages of heart failure or the New York Heart Association (NYHA) classification or a combination of both (20, 22) (Table 4 and Table 5). The NYHA classification is commonly used as a method for functional classification in patients with HF, in clinical practice and as an entry criterion and/or outcome measure in clinical trials. Although reflective of the natural course of HF, the patient's class assignment is performed by a clinician and relies on clinician's subjective interpretation of patient's functional capacity. Since concordance between cardiologists assigning NYHA classes can be as low as 54%, the validity of NYHA class as an outcome in clinical trials is disputed (32). This is a key reason why the economic model based on health states defined by NYHA is not presented in this submission.

Table 4: ACCF/AHA stages of chronic heart failure

Stage	Description
A	At high-risk for heart failure but without structural heart disease or symptoms of heart failure
B	Structural heart disease but without signs or symptoms of heart failure
C	Structural heart disease with prior or current symptoms of heart failure
D	Refractory heart failure requiring specialised interventions

Abbreviations: ACCF, American College of Cardiology Foundation; AHA, American Heart Association
Reference: (22)

Table 5: NYHA functional classification based on severity of symptoms and physical activity

Classification	Description
Class I	No limitation of physical activity. Ordinary physical activity does not cause symptoms of heart failure
Class II	Slight limitation of physical activity. Comfortable at rest, but ordinary physical activity results in symptoms of heart failure
Class III	Marked limitation of physical activity. Comfortable at rest, but less than ordinary activity causes symptoms of heart failure
Class IV	Unable to carry on any physical activity without symptoms of heart failure, or symptoms of heart failure at rest

Abbreviation: NYHA, New York Heart Association
Reference: (20)

B.1.3.1.2 Epidemiology

- Population growth, aging, and rising burden of diabetes and obesity are driving the increasing global prevalence of HF (9).
- In the UK, 920,000 people are estimated to live with HF and 200,000 people are newly diagnosed with HF each year (33-36).
- Heart failure patients have a higher proportion of comorbidities compared to cancer patients, with coronary heart disease (47.8%-61.1%) and hypertension (45.7%-54.6%) the most common comorbidities amongst men and women diagnosed with HF (17, 37-39).

Prevalence and incidence

Heart failure is a growing public health problem driven by increase in population size and age (9). Approximately 64.3 million people worldwide are estimated to have HF (36). Based on 2014 data, there are more than 920,000 people with HF in the UK (9)

and, of those, approximately 550,000 are on their General Practitioner's (GP) HF register (33). From 2002 to 2014, the prevalence of HF in the UK increased by 23% (9). In 2017, the prevalence of HF diagnosed in primary care was 2% and 5.9% among those aged 65-74 years and those older than 75 years respectively, with a higher estimated prevalence in men (7.5%) compared to woman (4.8%) in the over 75 year olds (33). The Quality Outcomes Framework estimates the prevalence of HFrEF to be 1.2% (2019/2020) (40). Recent retrospective studies indicate that HFrEF comprises 64% to 83% of the total HF caseload in the UK and is therefore more common than HFpEF or HFmEF (41, 42).

The number of newly diagnosed HF cases in the UK has increased by 12% in the period from 2002 to 2014 and there is no indication that the trend is slowing down (9). Between 176,000 and 200,000 people are newly diagnosed with HF each year in the UK, with the average age of diagnosis between 72 and 77 years (9, 33, 43). The age standardised incidence of HF is higher in men than in women (incidence rate ratio 1.52, 95% CI 1.50-1.54), with men also younger at diagnosis than women (mean age 74.0 years vs 79.4 years) (9). There has been a year-on-year increase in the incidence of HF since 2015. A recent UK RWE study (PULSE) reported that the incidence of diagnosed HF increased from 4.10 per 1000 person years in 2015 to 4.85 per 1000 person years in 2019 (43).

Prioritising the improvement of outcomes for HF patients is just as important as for other common conditions with a high burden of disease. The prevalence and incidence of HF in the UK is similar to the four most common causes of cancer combined (breast, prostate, lung and bowel) or COPD (9, 44). Between 2015 and 2017, Cancer Research UK reported the number of aggregated new cases for the four aforementioned cancers to be over 183,000 (45). Similarly, around 1.2 million people in the UK have COPD and approximately 115,000 people are newly diagnosed with COPD each year (44). The burden of HF is similar to cancer or COPD, indicating an urgent need to improve outcomes for HF patients at scale.

Comorbidities

The cardiorenal syndrome (CRS) encompasses a spectrum of disorders of the heart and kidneys whereby the physiological interdependence of the two organs leads to their simultaneous, accelerated decline in a negative feedback cycle (46). Metabolic disturbances associated with diabetes can also lead to the pathogenesis of the CRS by causing biochemical, functional and morphological abnormalities of the heart and kidney (47). HF patients therefore often suffer from renal or metabolic comorbidities due to the overlapping risk factors for these conditions (48). Nearly half of all HF patients have moderate to severe kidney dysfunction which increases the risk of hospitalisation or death compared to HF alone (10, 49, 50). Furthermore, nearly one-third have comorbid T2DM, also known to increase the risk of hospital admissions and cardiovascular (CV) death (10, 51). The onset of T2DM increases the risk of HF by two-fold in men and five-fold in women (52). Other comorbidities occurring with a high frequency in HF patients include ischaemic heart disease (IHD), hypertension, atrial fibrillation and diabetes (Table 6) (41).

Table 6: Causes and comorbidities of heart failure

Medical History	HFrEF (%)
IHD	46
Atrial fibrillation (from electrocardiogram (ECG))	41
Valve disease	27
Hypertension	52
Diabetes	34
COPD	18
Asthma	9

Abbreviations: COPD, chronic obstructive pulmonary disease; ECG, Electrocardiogram; IHD, Ischaemic heart disease
Reference: (41)

The burden of comorbidities is much higher for HF compared to other common conditions, such as cancer. A retrospective Scottish study conducted between 2002 and 2011 on adults with HF and four of the most common cancers showed that 94.5% of HF patients had comorbidities compared to 62%-80% of patients with a cancer

diagnosis (Table 7 and Table 8) (37). The data reported in the tables below further demonstrate the significant burden of HF disease to patients and the NHS.

Table 7: Baseline characteristics in men from Scotland with cancer, heart failure and comorbidities

	Prostate cancer	Lung cancer	Colorectal cancer	Bladder cancer	Heart failure
Cases, n	6,795	4,693	4,239	2,082	10,309
Heart failure, n (%)	95 (1.4%)	97 (2.1%)	81 (1.9%)	41 (2.0%)	-
Cancer, n (%)	-	-	-	-	226 (2.2%)
No comorbidity, n (%)	1,949 (28.7%)	1,116 (23.8%)	1,278 (30.1%)	499 (24.6%)	562 (5.5%)
Hypertension, n (%)	2,614 (38.5%)	1,515 (32.3%)	1,596 (37.7%)	801 (39.5%)	4,711 (45.7%)
Asthma, n (%)	491 (7.2%)	355 (7.6%)	286 (6.7%)	124 (6.1%)	788 (7.6%)
Coronary heart disease, n (%)	1,303 (19.2%)	1,091 (23.2%)	817 (19.3%)	488 (24.1%)	6,295 (61.1%)
Diabetes, n (%)	688 (10.1%)	562 (12.0%)	611 (14.4%)	314 (15.5%)	2,234 (21.7%)
COPD, n (%)	611 (9.0%)	1,241 (26.4%)	390 (9.2%)	237 (11.7%)	1,707 (16.6%)
Stroke or TIA, n (%)	321 (4.7%)	445 (9.5%)	245 (5.8%)	112 (5.5%)	754 (7.3%)
Previous MI, n (%)	657 (9.7%)	563 (12.0%)	442 (10.4%)	261 (12.9%)	4,448 (43.1%)
Chronic kidney disease, n (%)	550 (8.1%)	473 (10.1%)	381 (9.0%)	220 (10.8%)	1,560 (15.1%)
Atrial fibrillation, n (%)	238 (3.5%)	168 (3.6%)	162 (3.8%)	106 (5.2%)	552 (5.4%)
PVD, n (%)	388 (5.7%)	285 (6.1%)	250 (5.9%)	115 (5.7%)	2,519 (24.4%)

Abbreviations: COPD, chronic obstructive pulmonary disease; MI, myocardial infarction; PVD, peripheral vascular disease; TI, transient ischaemic attack
Reference: (37)

Table 8: Baseline characteristics in women from Scotland with cancer, heart failure and comorbidities

	Breast cancer	Colorectal cancer	Lung cancer	Ovarian cancer	Heart failure
Cases, n	10,760	3,610	3,859	1,234	9,131
Heart failure, n (%)	85 (0.8%)	43 (1.2%)	61 (1.6%)	15 (1.2%)	-
Cancer, n (%)	-	-	-	-	364 (4.0%)
No comorbidity, n (%)	4,115 (38.2%)	10,24 (28.4%)	769 (19.9%)	465 (37.7%)	500 (5.5%)
Hypertension, n (%)	3,259 (30.3%)	1,450 (40.2%)	1,451 (37.6%)	364 (29.5%)	4,984 (54.6%)
Asthma, n (%)	945 (8.8%)	296 (8.2%)	386 (10.0%)	95 (7.7%)	925 (10.1%)
Coronary heart disease, n (%)	839 (7.8%)	499 (13.8%)	718 (18.6%)	108 (8.8%)	4,367 (47.8%)
Diabetes, n (%)	786 (7.3%)	425 (11.8%)	421 (10.9%)	89 (7.2%)	1,708 (18.7%)
COPD, n (%)	583 (5.4%)	275 (7.6%)	1,118 (29.0%)	74 (6.0%)	1,455 (15.9%)
Stroke or TIA, n (%)	445 (4.1%)	237 (6.6%)	382 (9.9%)	58 (4.7%)	1,404 (15.4%)
Previous MI, n (%)	305 (2.8%)	207 (5.7%)	292 (7.6%)	48 (3.9%)	2,665 (29.2%)
Chronic kidney disease, n (%)	265 (2.5%)	179 (5.0%)	228 (5.9%)	37 (3.0%)	722 (7.9%)
Atrial fibrillation, n (%)	316 (2.9%)	158 (4.4%)	161 (4.2%)	25 (2.0%)	2,370 (26.0%)
PVD, n (%)	238 (2.2%)	130 (3.6%)	274 (7.1%)	30 (2.4%)	740 (8.1%)

Abbreviations: COPD, chronic obstructive pulmonary disease; MI, myocardial infarction; PVD, peripheral vascular disease; TI, transient ischaemic attack
Reference: (37)

Risk factors for disease

Risk factors associated with chronic heart failure can be modifiable (e.g. diet and exercise) or non-modifiable (e.g. age, gender, and comorbidities). Coronary heart disease, diabetes and age are strongly associated with increased risk of HF (10).

Hypertension, smoking, male gender, elevated body-mass index, diet and poor physical activity are also contributing to the pathogenesis of HF (53-55).

B.1.3.1.3 Disease burden

- Heart failure is a debilitating condition which substantially impacts patients' quality of life and disrupts work and lifestyle habits (20, 35, 56).
- In the UK, HF mortality ranges between 14.4%-26% at one year, 48.5%-68.1% at five years and 66%-75.5% at ten years (37, 39, 57-59). Comorbidities (CKD, diabetes, lung disease) and repeated HF hospitalisations are associated with even poorer long-term outcomes (37, 39, 57, 58).

Symptomatic burden

Heart failure patients experience debilitating symptoms including breathlessness, orthopnoea, paroxysmal nocturnal dyspnoea, reduced exercise tolerance, fatigue and ankle swelling (20, 56). Healthcare professionals describe HFrEF patients as highly symptomatic, with more severe breathlessness on exertion, with chronic fluid overload and pulmonary hypertension compared to other types of HF (35). The interdependencies within the CRM system lead to accelerated progression of CKD and HF and increase the symptomatic burden of HF patients (60). Furthermore, most HFrEF patients receive suboptimal daily doses of current treatments due to side effects and reduced tolerability associated with comorbidities (e.g. asthma and renal dysfunction) (61, 62). There is therefore a remaining need for new disease modifying HFrEF treatments.

Morbidity and mortality

The prognosis of HF remains poor, with the burden of HF in the UK similar in magnitude to that of the four most common cancers combined (9, 37, 39). Estimates for 1 and 5-year HF mortality in the UK are variable, but range between 14.4%-26% for 1-year and 48.5%-68.1% for 5-year post-diagnosis (37, 39, 57, 58). Notably, the 10-year mortality in the UK for HF was estimated at 75.5% in one UK study (57), with the mortality risk for HFrEF patients in another UK study reported to increase by an additional 37% from the reference control population (29%) equating to a 2.4 fold

excess loss of life in (59). Additionally, a further UK retrospective study of 241 people (200 with HFrEF and 41 with HFpEF) indicated that 32% of HFrEF patients died within 1-year of hospital admission (42).

Evidence suggests that HF mortality in the UK may be higher compared to a European RWE study and Norwegian and global study (63-65). These studies reported a 1-year mortality rate range of 6.4%-20% in chronic HF patients, a 5-year mortality of 45% in chronic HF patients, and a 1-year all-cause mortality and 5-year all-cause mortality range of 6.8%-8% and 33.5%-35% in HFrEF patients respectively (43, 63-65).

The overall prognosis of HF patients is exacerbated when patients have other comorbidities, such as CKD and diabetes (66-68). A UK national study reported that in patients who have both HF and CKD, hospitalisation and mortality rates increased by 11% and 17% respectively, compared to HF patients who do not have CKD (68). Furthermore, HF patients with diabetes showed a higher mortality rate of 34% compared to those without diabetes with a mortality rate of 22% from either a cardiovascular death or HF hospitalisation (67).

Healthcare system burden

Heart failure is the most common cause of hospitalisation in those over 65 years (37, 69-72). In general, it has been reported that repeat HF hospitalisations are known to be a strong predictor of mortality (72). A UK retrospective study reported a hospital readmission rate of 27% in the 12 months after discharge for HFrEF patients, and another study reported approximately 20%-30% of HF patients having a hospital readmission within 30 days, rising to 50% at the 6-month time point (42, 73). Higher rates of hospitalisation for heart failure (HHF) are observed in HF patients with diabetes, where the readmission rate is nearly double compared to those without diabetes (74-76). Furthermore, current HF treatments that reach suboptimal dosage also contribute to the increased risk of HF-related hospitalisations and mortality (61, 77). Therefore, there is an unmet need for new HFrEF treatments to lower hospitalisation rates and reduce mortality.

The morbidity and mortality burden of HF in the UK impedes the provision of holistic care, a key priority in the NHS Long-Term Plan. With every subsequent hospitalisation for HF, the risk of having an unplanned death increases and the opportunity to die with dignity reduces. The choice on how to die is taken away from the patient. One study reported that 71.5% of HF patients who had an unplanned readmission or death within 30 days of index discharge, experienced at least one emergency attendance in the 6 months prior, which was higher than those patients who did not have a readmission (83, 84).

B.1.3.1.4 Economic burden

- In the UK in 2012, the direct and indirect costs of HF amounted to £2.0 billion and £888 million, respectively (23, 78-80).
- HFrEF patients who have T2DM have higher hospitalisation rates and longer length of stays, and the burden is substantial amongst HF patients with CRM-related comorbidities (81, 82).

There is a substantial economic burden of HF in the UK, where it is estimated to annually account for 2% of the National Health Service (NHS) budget, with 60%-70% of the costs related to hospitalisations (23, 80). Heart failure patients accounted for 1 million inpatient bed days (representing 2% of all NHS inpatient bed days and 5% of all emergency medical hospital admissions) with an average length of stay of 6 to 9 days and a three month readmission rate of 25% (23). Additionally, patients with HFrEF and T2DM have higher hospitalisation rates and longer length of stays compared to those without T2DM (81). Notably, in HF patients the burden of CRM-related conditions is substantial in terms of the cost burden and all-cause hospital admissions and this is further amplified in the T2DM population (82). In 2012, it was estimated that the direct and indirect costs of HF amounted to ~£2.0 billion and £888 million, respectively (23, 78, 79).

Inpatient care or critical care account for more than 90% of health care costs during the last three months of a HF patient's life (85). Informal care costs can also rise with increasing hospitalisations and caregivers burden also significantly impacts both leisure time and productivity as caregiving responsibilities result in 28 hours per week

of time commitment (86-88). Furthermore, there is a high proportion of patients in the UK with underdiagnosed or undertreated chronic HF, which is associated with high costs and increased risk of death (89). The substantial economic burden indicates a need to reduce HF hospitalisation costs given that this is a dominating contributor towards total HF costs in the UK.

B.1.3.1.5 Humanistic burden

- Heart failure has a substantial impact on patient's and carer's quality of life affecting their social, emotional and psychological well-being (35).
- Patients with HFrEF are known to have significantly lower QoL than patients with HFmEF and HFpEF (90).

Heart failure has a significant impact on patient's physical and emotional well-being. Often, the Kansas City Cardiomyopathy Questionnaire (KCCQ) is used to quantify humanistic burden in HF patients (on a scale between 0 and 100), since it is an established patient-reported measure of health status, is valid, reliable, sensitive, specific and responsive to HF quality of life (91, 92). The KCCQ domains quantify the patient's perception of their health status including HF symptoms (frequency and burden), physical and social limitation and QoL, with higher scores indicating better health status, lower symptom burden and better HRQoL (91, 92). This compares to the NYHA classification which is a physician's interpretation of patient's symptoms, and can often lead to biased assessments, whereas the KCCQ is a more robust, patient-centric questionnaire and is likely to be used more commonly in clinical trials given the sound KCCQ psychometric properties (32, 91, 92).

Physical well-being of HF patients was reported in several UK studies, where patients experienced a range of symptoms including breathlessness, reduced sleep quality, frailty, cognitive/psychomotor impairment, respiratory symptoms and chest pain (20, 56, 93). One UK study reported a continuous quality of life difference in chronic HF patients compared to those without HF, where on average a 16% reduction in physical activity was observed (93). The impact of heart failure on emotional well-being is significant. Patients have often reported feeling overwhelmed, frustrated, limited and

worried, particularly around the caring for their children/spouses and the impact it has on their self-confidence (35).

Equally, carer's health as a result of carer's responsibilities were also significantly impacted by stress (35%), moderate to severe anxiety/depression (32%), emotional strain (33%), physical (33%) or mental (31%) tiredness and pain/discomfort (29%) (88). The substantial reduction in patient's physical and emotional well-being are even associated with a higher risk of mortality (56, 90, 93-97). Quality of life and risk of mortality are further impacted when patients are hospitalised with HF (97).

B.1.3.1 Clinical pathway of care

B.1.3.1.1 Current standard of care

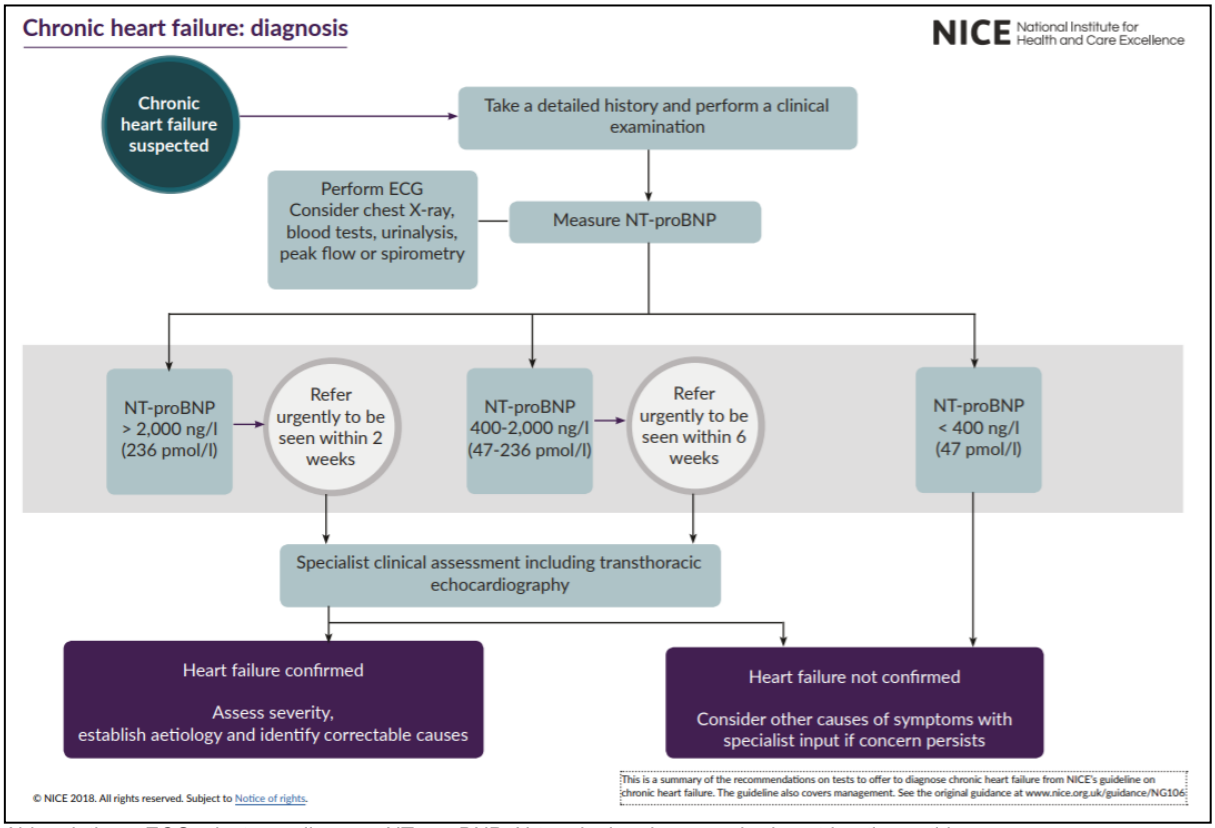
- The National Clinical Guideline 106 (NG106) recommends a sequential approach to management of chronic HF, ACEi or ARBs in combination with a BB comprising the first line of treatment that can be prescribed and monitored in primary care or by HF specialists (34).
- Specialist treatments for HFrEF patients who remain symptomatic on standard care (sacubitril valsartan, ivabradine, hydralazine and nitrate, and digoxin) serve a limited population in clinical practice (34).
- Although NG106 aims to standardise provision of HF services, the quality of HF care varies with regards to diagnosis, follow-up, tools to quantify HF symptom severity, patients' well-being and rehabilitation programmes (23, 35, 39, 80).
- In clinical practice, a large proportion of patients do not receive the guideline-recommended doses of standard treatment because of comorbidities and side effects. In the UK, the average doses of ACEi/ARB and BB received by HF patients comprise just 48% and 40% of the recommended dose, respectively, which are associated with greater risk of death and/or HHF (61, 62, 77).
- In the UK, HF care pathways are variable and optimisation of current treatments can be considerably longer than recommended (20, 34). Restrictions on prescribing of SGLT2is risk exacerbating this trend and widening the gap in health inequalities.
- Continued use of evidence-based guidelines in conjunction with long-term prevention programmes, multidisciplinary care (e.g. GPs with specialist interest and nursing specialists in primary care) and interventions to control cardiometabolic risk factors are required to improve HF outcomes (23, 39, 80).

NICE clinical treatment pathway

The diagnosis of HF is multifactorial and encompasses detailed clinical history, physical examinations, electrocardiograms (ECG), stress tests, chest x-rays, coronary angiograms, cardiac computerized tomography (CT) scans, magnetic resonance imaging (MRI), myocardial biopsies and laboratory tests. Given the uncertainties that are intrinsic to a clear diagnosis of HF on physical examination alone, and the outcome for patients left undiagnosed, the NICE and ESC guidelines recommend testing of serum N-terminal pro-B-type natriuretic peptide (NT-pro-BNP) in people with

suspected heart failure as essential diagnostic tool (Figure 2) (20, 34). The NT-pro-BNP level however cannot differentiate between HFrEF and HFpEF. Transthoracic echocardiography is required for confirmatory diagnosis and to inform classification of HF, which in turn, guides the management of the condition (20, 34).

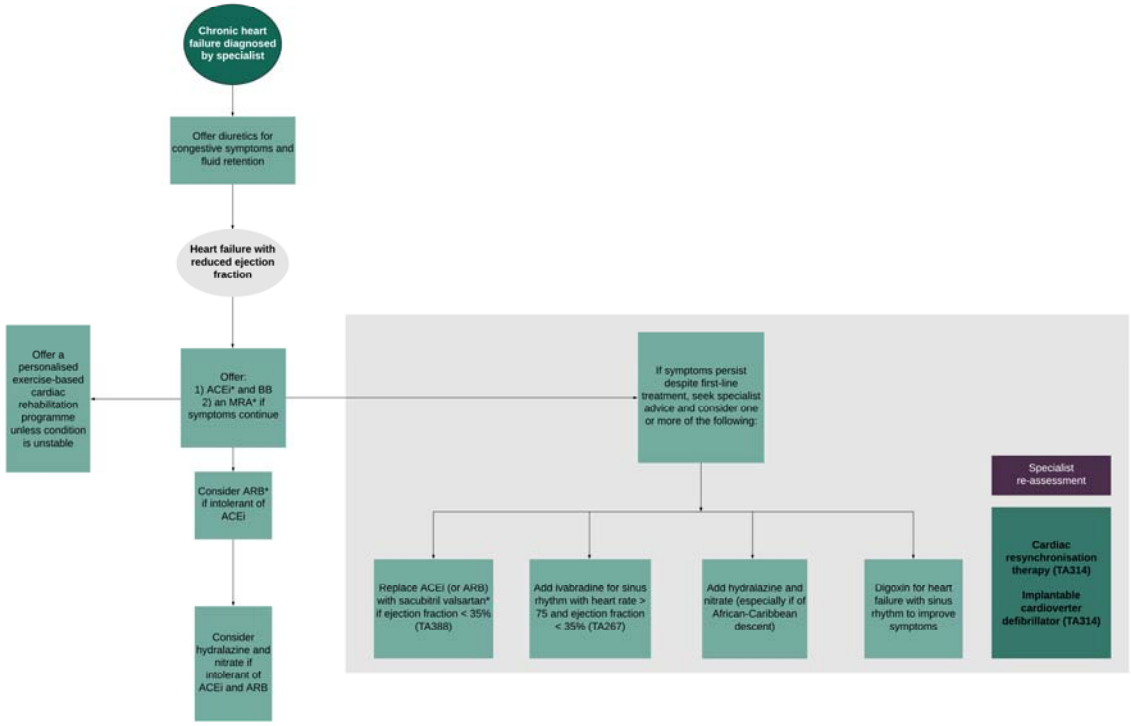
Figure 2. Chronic heart failure diagnostic pathway



Abbreviations: ECG, electrocardiogram, NT-pro-BNP, N-terminal prohormone brain natriuretic peptide.
Source: NICE guideline NG106 (34)

Following HFrEF diagnosis, the main goals of the treatment are improvement in prognosis, physical functioning, and symptom burden. The NICE guideline for HFrEF management recommends a sequential approach, with available pharmacotherapies divided into the first line and the specialist treatments (Figure 3) (34).

Figure 3. Chronic heart failure NICE treatment pathway



*Measure serum sodium and potassium and assess renal function before and after starting and after each dose increment. If estimated glomerular filtration rate (eGFR) is 30 to 45 ml/min/1.73m^2, consider lower doses or slower titration of ACEi or ARBs, MRAs, sacubitril valsartan and digoxin.
Abbreviations: ACEi angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor-blocker; BB, beta-blocker; MRA, mineralocorticoid receptor antagonist.
Source: Adapted from NICE guideline NG106, 2018 (34)

First line treatments are the most relevant comparators to consider for this appraisal. The first line treatment comprises an angiotensin-converting enzyme inhibitor (ACEi) (or angiotensin receptor-blocker (ARB) if intolerant to ACEi) and a beta-blocker (BB) (34). Diuretics are also used routinely in the first line treatment to provide symptomatic relief, particularly in the presence of oedema, but without direct evidence of survival benefit. If symptoms continue, mineralocorticoid receptor antagonists (MRA) can be added to the ACEi or ARB if there is no evidence of hyperkalemia (34). Health care professionals are advised to reach the target dose for each drug class before prescribing another drug class. Although general practitioners can prescribe and titrate ACEi/ARB to their maximum tolerated dose (MTD), in clinical practice, the setting for initiation and up-titration of first line treatments varies depending on geographic region and advice from a HF specialist is often sought due to limited confidence among

general practitioners' (34, 35). A HF specialist could be a cardiologist, HF nurse or pharmacist. The MRA initiation is commonly carried out by a HF specialist owing to the risk of hyperkalemia, hypertension, or renal impairment (34, 35).

Specialist treatments are considered when symptoms persist after dose optimisation of standard of care (SoC) therapy with ACEi/ARB, BB and/or MRA combination, and require supervision of a HF specialist with access to a multidisciplinary team during initiation and optimisation. Figure 3 outlines the available specialist treatments, which include:

- Sacubitril valsartan (Entresto®) as an alternative to ACEi or ARB in patients with continuing NYHA class II-IV symptoms and LVEF ≤ 35%
- Addition of ivabradine to the SoC for patients in sinus rhythm with a heart rate ≥ 75 beats per minute and LVEF ≤ 35%
- Addition of hydralazine and nitrate especially in patients of African-Caribbean descent with moderate to severe HF, or in patients who can tolerate neither an ACEi nor an ARB
- Digoxin in patients with worsening or severe HFrEF with sinus rhythm and reduced renal function.

For patients who have HFrEF and CKD (eGFR <45ml/min/1.73m^2), lower doses and/or slower titration of ACEi or ARBs, MRAs and digoxin should be considered. The NICE recommendations for HFrEF diagnosis and treatment are consistent with guidelines of the European Society of Cardiology (ESC) (20, 34).

Clinical practice and heart failure services

Evidence suggests that the treatments for HF patients recommended in the clinical guidelines have been widely adopted in the UK clinical practice (20, 34). A high proportion of HFrEF patients in the UK have been reported to receive ACEi/ARB (70-85%) and BB (89%) in the National Heart Failure Audit, and other retrospective studies have confirmed this trend (ACEi/ARB, 81.0%; BB, 73.1%) (41, 43, 98). The treatment rates with MRA (27.0%), digoxin (15.3%), ivabradine (3.0%) and hydralazine/nitrate (0.6%) were considerably lower, in line with their more restrictive eligibility criteria and contraindications (MRA, ivabradine, hydralazine/nitrate) and less certainty regarding

health benefits (digoxin) (43). The low uptake of ivabradine, hydralazine/nitrate and digoxin, which are often used to treat comorbid conditions common in HF population (stable angina pectoris, hypertension and dysrhythmias, respectively) rather than HF per se, suggests that these treatments are not relevant for the broad population of HFrEF patients in which empagliflozin in indicated (43). This is consistent with the NICE technology appraisal for dapagliflozin (TA679), where clinical experts fed back that ivabradine, hydralazine and nitrate, and digoxin are used in a specialist setting only and were not relevant comparators for dapagliflozin (6).

Current first line SoC treatments are non-curative and often require gradual up-titration to the achieve guideline-recommended dose associated with the optimal clinical benefit (34). Health care professionals are advised to up-titrate each drug class to its recommended dose before prescribing another drug class (34). However, in clinical practice across Europe, most HFrEF patients receive suboptimal daily doses of ACEi/ARBs and BBs: after up-titration in patients with new onset or worsening symptoms of existing HFrEF, only 22% reached the recommended dose of ACEi/ARB and 12% reached the target BB dose (61, 62, 77). Advanced age, low heart rate and comorbidities such as asthma and hypotension were the main predictors of lower than target dose of BB, while female gender, lower BMI and renal impairment were the main reasons behind suboptimal ACEi/ARB doses (61, 62, 77). Crucially, reaching less than 50% of the target ACEi/ARB and/or BB dose is associated with a higher risk of mortality and combined endpoint of death or HHF compared to those reaching 100% of the recommended dose (ACEi dose 1-49%: combined HR 1.23, 95% CI 1.09–1.36; BB dose 1-49%: combined HR, 1.27, 95% CI 1.15-1.39). Furthermore, many patients receiving suboptimal doses of SoC who remain symptomatic are ineligible for MRAs or sacubitril valsartan due to declining renal function and/or CKD, which limits their treatment options (35).

There are also inconsistencies between the guidelines and clinical practice in HF service settings (e.g. hospital-based, community-based, hospital and community-based or hospital with community work) (34, 39). Optimal management of chronic HF requires optimisation of pharmacological treatment, nursing support and treatment of comorbidities and should be delivered by a multidisciplinary team (MDT) (34, 39).

Ideally, the MDT should integrate the community and hospital care, and consist of a consultant cardiologist, a specialist heart failure nurse, GP, pharmacist, physiotherapist, palliative care, psychologist, occupational therapist and administrators (38). The objective of the heart failure MDT is to ensure patients receive continuous, individualised care, which is responsive to their changing needs when transitioning between primary and secondary care (38).

In the UK clinical practice, however, the structure and provision of HF care varies considerably by geographic location and is not always consistent with the current guidelines (34, 39). Variation has been observed in access to natriuretic peptide testing for diagnosis and monitoring, use of validated tools to quantify the severity of symptoms and monitor quality of life and in the availability and uptake of rehabilitation programmes (39, 99). Although the availability of echocardiography services for HF diagnosis is generally high, limited knowledge of the interpretation of results in primary care results in frequent referral to specialist HF clinics and delays in diagnosis. Importantly, the variability in service provision has an adverse impact on the follow-up and monitoring of pharmacological management (39, 99). A substantial proportion of GPs do not routinely initiate diuretics (23%), ACEi (22%) or BBs (38%) for HF with left ventricular systolic dysfunction for reasons that include challenge of managing renal comorbidities and side effects of the HF polypharmacy, with the added burden of dose titration and monitoring (93).

A recent Clinical Practice Research Datalink (CPRD) study identified that eight out of ten HF cases in England are diagnosed after emergency hospital admission for acute HF symptoms, while fewer than a quarter of those with HF symptoms recorded in primary care follow the recommended clinical pathway for investigation and specialist referral (100, 101). This divergence of practice from the guideline has serious implications for HF outcomes given a significantly higher risk of all-cause mortality in the hospital-diagnosed compared with the community-diagnosed HF patients (HR 1·55, 95% CI, 1·53 - 1·58) (102). Furthermore, only half of HF patients in the UK are reviewed within the target two weeks from the initiation of change of medication due to capacity and staffing limitations (39). Hence, dose optimisation of current treatments

can take six months or longer (NICE recommends six weeks), resulting in a delayed mortality and morbidity benefit from the SoC (101, 103). Significant unmet need in chronic HF therefore stems from suboptimal implementation of the clinical guideline, in particular in relation to access to a HF consultant and specialist nurses (35, 39).
COVID-19 has severely impacted delivery of care and patients are less likely to seek medical care for any HF symptoms they experience (B.1.4), (15). COVID-19 has exacerbated pre-existing health inequalities, as patients in a lower socioeconomic group were less likely to seek medical attention in secondary care before the pandemic. The impact is significant. HF is also a risk factor for COVID-19 (104, 105). Additionally, HF patients with a lower socio-economic status were already more likely to have worse CV outcomes than those with a higher socio-economic status. Patients with chronic heart failure were 17% more likely to die of COVID-19 than those who did not(106).

B.1.3.1.2 Unmet need

- Patients with chronic HF continue to experience high mortality and morbidity, high symptom burden, reduced functional capacity and poor QoL.
- Most HFrEF patients are on suboptimal doses of ACEis, ARBs and BBs due to comorbidities and tolerability issues, as well as delays in dosing optimisation in secondary care, which contributes to their poor prognosis (66, 81).
- Concurrent improvements in the timely diagnosis of HFrEF and the existing treatment paradigm (i.e. simultaneous sequencing and dose optimisation of available therapies) remain the key priorities in delivery of HF care.
- Improving the existing treatment paradigm will take time and there remains a significant residual risk associated with HFrEF. There is an urgent need for disease-modifying therapies that are available across primary and secondary care that have an immediate impact on patient prognosis and QoL without dose-limiting side effects. This will be important for post COVID-19 recovery.

As mentioned in B.1.3.1.2, chronic HF affects around 1 million people in the UK, of which up to two-thirds are estimated to have HFrEF (9, 107). More than half of all HF patients die within five years from diagnosis, with cardiovascular disease as the most

common cause of death (57). The prevalence and severity of HF and HFrEF at onset are expected to increase in the next decade due to an aging population, overall population growth and the rising risk of diabetes mellitus, obesity and CKD, the known risk factors for HF (10).

The mortality and morbidity of HF remain high due to a number of factors, including late diagnosis that most often occurs after emergency admission for acute HF symptoms, delays in optimisation of pharmacological therapy, widening socio-economic inequalities and variation in access to HF care depending on location (9, 39, 100, 103). Further, with each subsequent hospitalisation, the risk of an unplanned death increases, reducing the opportunity for a patient to choose how they want to die (B.1.3.1.3) (9, 39, 100, 103). Significant divergence of clinical practice from the NICE guideline leads to poorer outcomes, as patients diagnosed after hospitalisation have significantly higher all-cause mortality compared to the community-diagnosed group, and those whose treatment pathway does not, at least partially, adhere to the guideline, have an increased risk of HHF (101, 102). In 2018/19, there were more than 100,000 hospital admissions for HF in the UK, an increase of almost a third compared to 2013/14 (102, 108).

Although the existing HFrEF treatments, ACEi/ARBs, BBs, MRA and sacubitril valsartan, are associated with reduced mortality, many patients do not receive the recommended doses of treatment due to age and comorbidities, particularly hypotension, renal impairment and asthma, and are at a greater risk of death and/or HHF than those receiving the target doses of ACEi/ARB and BB (61, 62, 77). The failure to achieve recommended dosing is also partly attributable to adverse events associated with current therapies (e.g. hypotension, renal function decline, and hyperkalemia) as well as the duration of time required to up-titrate to MTD (≥ 6 months), with patients less likely to adhere to lengthy treatment process without the immediate clinical improvement (61, 62, 77, 103). Furthermore, the safety profile of sacubitril valsartan limits its dose or precludes its prescribing in a large subset of HFrEF patients with hypotension, electrolyte imbalance or comorbid CKD (40). Other specialist HFrEF treatments, ivabradine, hydralazine with nitrate, and digoxin are

rarely prescribed in clinical practice and are primarily used to treat comorbid conditions common in HF population such as angina and arrythmias (6, 43).

Concurrent improvements in the timely diagnosis of HFrEF and the existing treatment paradigm (i.e. simultaneous sequencing and dose optimisation of available therapies) remain the key priorities in delivery of HF care. Greater involvement of GPs with specialist interest (GPwSI) and specialist nurses in primary care, who are knowledgeable in diagnosis and management of HF could support these improvements. The GPwSI could upskill GPs and specialist nurses to facilitate earlier diagnosis, initiation and monitoring of HF treatments that reduce risk of emergency hospitalisation and thus reduce the healthcare burden of HF (111, 112).

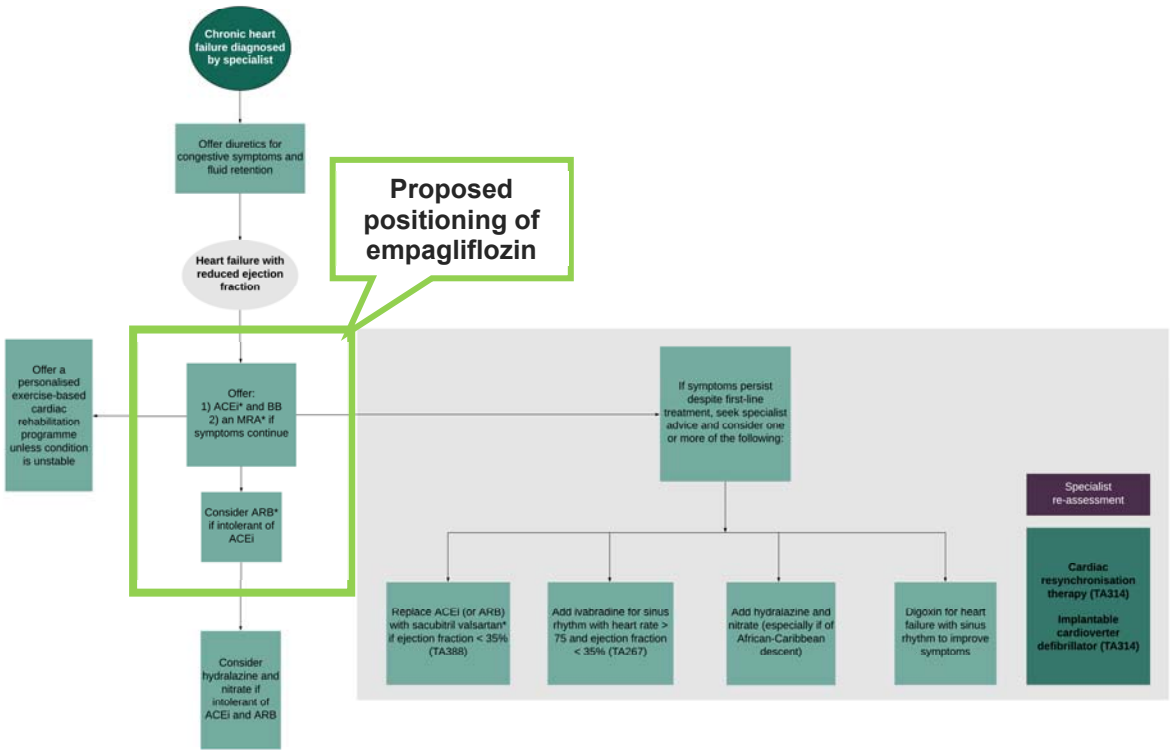
Optimisation of the existing treatment paradigm will take time and there remains a significant residual risk associated with HFrEF. Thus, there is an urgent need for disease-modifying therapies that are available across primary and secondary care that have an immediate impact on patient prognosis and QoL without dose-limiting side effects. This will be important for post COVID-19 recovery.

B.1.3.1.3 Positioning of empagliflozin in the UK treatment pathway

- Empagliflozin as an add-on to SoC leads to a significant reduction in the risk of CV death or HHF, and a sustained improvement in renal outcomes and HRQoL compared to SoC alone (109).
- As empagliflozin requires no dose adjustment, it could be initiated in primary care, reducing capacity burden in secondary care (112-114); a major cause of delay in optimising patients' SoC treatment according to NG106.
- With broad prescribing of empagliflozin across primary and secondary care, there is an opportunity to maximise outcomes for HF patients immediately.
- Broad prescribing of empagliflozin across primary and secondary care reduces the risk of widening the gap in health inequalities seen in HF patients as a result of COVID-19 (B.1.4).

Based on the population studied in the pivotal phase III study EMPEROR-Reduced, the optimal positioning for empagliflozin in the NICE pathway is as an add-on to ACEi or ARBs plus BB, and/or MRA therapy for HFrEF patients with or without comorbidities, who continue to be symptomatic while receiving stable, but not necessarily optimised doses of SoC (Figure 4).

Figure 4: Empagliflozin in the HFrEF treatment pathway



*Measure serum sodium and potassium, and assess renal function before and after starting and after each dose increment. If eGFR is 30 to 45 ml/min/1.73m^2, consider lower doses or slower titration of ACEi or ARBs, MRAs, sacubitril valsartan and digoxin.
Abbreviations: ACEi angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor-blocker; BB, beta-blocker; MRA, mineralocorticoid receptor antagonist
Source: Adapted from NICE guideline NG106, 2018 (34)
The preferred positioning for empagliflozin (in light green box) is in primary care in recently diagnosed symptomatic patients who receive SoC (e.g. ACE, BB, ARB), but not necessarily with optimised dosing

In this positioning, the relevant comparators for empagliflozin are:

- Individualised standard care defined as
  - ACEi in combination with a BB, and/or MRA
  - ARB in combination with a BB, and/or MRA
  - Sacubitril valsartan in combination with BB, and/or MRA

This is consistent with the perspective of UK clinicians BI have consulted. Further, a majority of HFrEF patients in the UK receive at least one of these products (Table 2).

Evidence for the comparison against individualised standard care stems from EMPEROR-Reduced trial, which showed empagliflozin to be an effective add-on to first line therapies, regardless of the dose used for these therapies (B.2) (109, 110). The composite primary outcome in EMPEROR-Reduced showed that 19.4% of patients receiving empagliflozin plus SoC vs. 24.7% receiving SoC alone experienced either a HHF or CV death event (ITT, HR 0.75, 95%CI, 0.65 - 0.86, p<0.001) (109). Empagliflozin plus SoC also demonstrated an improvement in kidney outcomes. A composite renal outcome (chronic dialysis or renal transplantation or a profound, sustained reduction in the eGFR) occurred in 30 patients (1.6%) in the empagliflozin plus SoC and in 58 patients (3.1%) in the SoC alone group (ITT, HR 0.50; 95% CI, 0.32 - 0.77). Empagliflozin has an established safety profile, requires no dose adjustment and thus no additional clinical time is needed to optimise a patient's treatment (109). Empagliflzoin has demonstrated improvement in HF-related outcomes across a broad range of populations including the presence or absence of T2DM and/or CKD, use of sacubitril valsartan at empagliflozin initiation, baseline health status as measured by KCCQ, younger or older age (109, 111-113).

Comparative analyses of empagliflozin vs dapaglflozin are provided in B.2.8 upon the request of the ERG and NICE Technical team; however these are secondary. It is clear from Table 2 that prescribing of dapagliflozin is not SoC at the time of submission, there is minimal scope for displacement, and hence we do not regard this comparator of primary relevance to the decision problem.

A NICE recommendation for empagliflozin in HFrEF will likely have a positive impact on the existing pathway. General practitioners' experience with prescribing SGLT2i in T2DM and there being no requirement for dose adjustment should facilitate initiation of empagliflozin in HFrEF patients within primary care (114). A recently proposed algorithm for sequencing of HFrEF treatments suggests early initiation of a BB and an SGLT2i with stable HFrEF and normal fluid balance, in parallel with referral for specialist treatments, could maximise prevention of deaths and HHF by avoiding

delays caused by sequential prescribing after dose optimisation of previous regimen ((103, 115)).

B.1.4 Equality considerations

Socio-economic inequalities in CV disease present a major and persistent UK public health challenge. The UK-based population studies demonstrate that socio-economic deprivation is a strong risk factor for the development of HF and adverse HF outcomes (9, 10). Individuals in the lowest socio-economic group are 1.61 times more likely to experience incident HF than the most affluent individuals and do so, on average, at a 3.5 years younger age with a greater comorbidity burden at time of HF symptom onset (9). Furthermore, the socio-economic status is associated with a diverging trend in HF outcomes in England, whereby patients from the most deprived group have a significantly higher risk of all-cause (HR, 1.17; 95% CI,1.14-1.21) and CV mortality (HR, 1.18; 95% CI, 1.14-1.23) than the most affluent ones (102).

Since the early 2000s, the socio-economic gradient in HF incidence and outcomes has been widening:

- The mean age at diagnosis increased by 2.45 years (95% CI, 1.58-3.32) among the most affluent but tended to decrease among the most deprived (9).
- The annual risk in HHF has increased by 1.6% (95% CI, 0.6-2.6) for the most deprived compared to a stable risk for the most affluent group (102).

The inequality in access to specialist care in the UK may be one of the drivers of the observed trends in HF. After controlling for need, richer individuals tend to consume more public and private specialist visits, but not family physician visits, than those from a lower socio-economic class, and experience significantly shorter waiting times for a coronary revascularisation procedure at the same public hospital (11, 116). The prominent role of a secondary care specialist in all aspects of HF care (including diagnosis, management, and initiation of new medicines) that can only be accessed upon referral from a GP, could therefore be contributing to the observed socio-economic disparities in clinical characteristics and outcomes of HF (34).

The choice of setting for empagliflozin initiation in primary care or under specialist supervision is thus a highly pertinent public health issue. Broad prescribing of empagliflozin across primary and secondary care can support the reduction in disparity in access to HF care across socio-economic groups within the UK, given that empagliflozin significantly improves CV and renal outcomes of HFrEF patients in an early, sustained manner and prevents hospitalisations for HF compared to SoC (109, 113). It is the only HFrEF treatment that can simultaneously provide cardiac, renal and glucose-lowering benefits to the large subset of patients with comorbid diabetes and/or severe renal impairment (eGFR 20 to 30 mL/min/1.73m^2), which are more likely to coexist in the most deprived patients (10). Limiting initiation of empagliflozin to secondary care specialists could lead to a delayed and/or lower uptake of empagliflozin among the most socioeconomically disadvantaged groups as they consume fewer specialist visits and present to health care providers at a later stage of illness (116). Delayed exposure to the benefits of SGLT2 inhibition may in turn widen the existing divide in HF outcomes between socio-economic classes in England.

The current COVID-19 pandemic may further reinforce this trend through significant disruption in the provision of all types of cardiology services including outpatient and community HF services (117). Patterns of past care suggest that the elderly and those living in deprived areas are most likely to be disproportionately affected by increased waiting times for cardiology appointments (11, 118). With a condition that has a 1-year mortality of approximately 24% and is the leading cause of hospital readmissions, a long wait for a HF specialist appointment may have grave consequences for socioeconomically disadvantaged HF patients in England (57).

In this health technology appraisal, equity of access to empagliflozin among HFrEF patients from all socio-demographic groups is an important consideration. A recommendation by NICE that facilitates broad prescribing of empagliflozin across primary and secondary care and its classification as “green” on local/regional formularies would support this objective. This in turn will support the overarching goal of reducing inequity in access to care for HFrEF patients, in line with NICE’s Social Value Judgments, pillar 3 of NICE’s new 5-year strategy (13, 14) and the conclusions from the Marmot COVID-19 Build Back Fairer review(15).

B.2 Clinical effectiveness

- EMPEROR-Reduced was an event-driven, double-blind RCT which enrolled 3730 patients with moderate to severe HFrEF (LVEF≤40%, NYHA II-IV) randomly assigned to receive empagliflozin (N=1863) or placebo (N=1867) in addition to all appropriate treatments for HF
- After a median follow-up of 16 months, empagliflozin significantly reduced the risk of death from CV causes or hospitalisation for HF compared to placebo (HR, 0.75; 95%CI,0.65-0.86; p<0.001)
- Empagliflozin was superior to placebo with respect to key secondary endpoints:
    - It led to a significant reduction in the total number of hospitalisations for HF (HR,0.70; 95% CI, 0.58-0.85; p<0.001) vs placebo
    - The rate of the decline in the estimated GFR was slower in empagliflozin group compared to placebo group over the duration of the double-blind treatment period (between group difference, 1.73ml/min/1.73m^2 per year; 95%CI, 1.10-2.37; p<0.001)
- Empagliflozin was also superior to placebo in several exploratory endpoints:
    - Reduced risk of an adverse renal outcome (composite of chronic dialysis, renal transplantation or a profound, sustained reduction in eGFR) (HR, 0.50; 95%CI, 0.32-0.77)
    - Reduced occurrence of all-cause hospitalisation (HR, 0.85; 95%CI, 0.75-0.95)
    - Improvement in QoL score on KCCQ at 52 weeks (HR, 1.7; 95%CI, 0.5-3.0)
- The impact of empagliflozin on exploratory endpoints CV mortality and all-cause mortality was numerically favourable but not statistically significant (HR, 0.92; 95%CI, 0.75 - 1.12) and HR, 0.92; 95% CI, 0.77-1.10)
- The HF and renal benefits of empagliflozin were consistent across subgroups of HFrEF patients defined by age, gender, use of neprilysin inhibitor, presence or absence of diabetes and/or CKD
- In diabetic patients, the effect of empagliflozin effectively negated the deleterious effect of diabetes on the risk of adverse CV and renal outcomes
- Empagliflozin improves CV and renal outcomes of HFrEF patients including those with an estimated eGFR of 20 ml/min/1.73m^2

B.2.1 Identification and selection of relevant studies

A systematic literature review (SLR) was conducted to identify from the published literature randomised controlled trial (RCT) evidence on the efficacy and safety of empagliflozin and relevant comparators in patients with chronic HF (NYHA class II-IV) with reduced LVEF. Full details of the process and methods used to identify and select the clinical evidence relevant to the technology being appraised are described in Appendix D.

Searches of Medline® and Embase® (via Embase.com), Medline in-Process® (via PubMed.com) and the Cochrane Library were performed on 14 May 2020, and subsequently updated on 8 October 2020. The search of electronic databases was supplemented with a desk search of conference proceedings, last conducted on 12 October 2020.

The eligible studies encompassed all RCTs evaluating efficacy of pharmacological interventions used in the treatment of adults (age ≥ 18 years) with chronic HFrEF. The search strategy was designed to be broad and to encompass all interventions that currently comprise the SoC, as well as recently approved interventions and investigational agents for the management of chronic HF (eligibility criteria are shown in Table 6 in Appendix D). All studies meeting the pre-specified PICOS eligibility criteria were retained; only studies of empagliflozin and interventions likely to be compared to empagliflozin were extracted in full [i.e. studies of SGLT2 inhibitors and ARNi (sacubitril valsartan)].

The EMPEROR-Reduced trial compared empagliflozin with the NHS SoC, and is therefore the primary source of clinical evidence in the economic model. The SLR identified six citations describing the design and outcomes of the pivotal trial of empagliflozin in HFrEF, EMPEROR-Reduced (109, 119-123). A full list of studies that were included and excluded during the SLR is provided in Appendix D.

B.2.1.1 Clinical trials with empagliflozin 10 mg (Jardiance®)

Empagliflozin is being investigated in the EMPOWER clinical trial programme. The most comprehensive development programme for an SGL2T inhibitor to date, EMPOWER is comprised of nine clinical trials and a real world evidence study that have been designed to evaluate the impact of empagliflozin on cardiovascular and renal outcomes of patients across the spectrum of CRM disorders (Table 9). Furthermore, the aim of the programme is to advance the scientific understanding of the pathophysiology of cardiorenal interactions and enable a holistic management of the interconnected CRM organ system.

Table 9. Overview of the studies comprising the EMPOWER clinical trial programme for empagliflozin

Study name	Study identifier	Main objective	Status	Relevant for this appraisal & reason
EMPEROR-Reduced	NCT03057977 (124)	Efficacy & safety of empagliflozin in prevention of CV death and hospitalisation due to HF in adults with chronic HFrEF with or without T2DM	Completed	Yes; meets the PICO criteria as defined in the decision problem
EMPEROR-Preserved	NCT03057951 (125)	Efficacy & safety of empagliflozin in prevention of CV death and hospitalisation due to HF in adults with chronic HFpEF with or without T2DM	Ongoing	No; population not relevant for the decision problem
EMPERIAL-Reduced	NCT03448419 (126)	Effect of empagliflozin on functional ability and PROs in adults with chronic HFrEF with or without T2DM	Completed	No; primary outcome not relevant for the decision problem; QoL secondary endpoint measured using PROs not recommended by the NICE reference case (85)

Study name	Study identifier	Main objective	Status	Relevant for this appraisal & reason
EMPERIAL-Preserved	NCT03448406 (127)	Effect of empagliflozin on functional ability and PROs in adults with chronic HFpEF with or without T2DM	Completed	No; population not relevant for the decision problem
EMPA-REG OUTCOME	NCT01131676 (128)	Efficacy & safety of empagliflozin in prevention of major adverse CV events, including CV death, in adults with T2DM and established CV disease	Completed	No; population not relevant for the decision problem
EMPULSE	NCT04157751 (129)	Efficacy of empagliflozin in improving clinical and PRO outcomes in adults hospitalised for acute HF	Ongoing	No; population not relevant for the decision problem
EMPA-KIDNEY	NCT03594110 (130)	Effect of empagliflozin on progression of kidney disease and the occurrence of CV death in patients with pre-existing CKD	Ongoing	No; population not relevant for the decision problem
EMPA-VISION	NCT03332212 (131)	Effects on cardiac physiology and metabolism in patients with HF	Completed	No; the study outcomes not relevant for the decision problem

Study name	Study identifier	Main objective	Status	Relevant for this appraisal & reason
EMPACT-MI	NCT04509674 (132)	Efficacy of empagliflozin in improving outcomes and preventing HF in adults hospitalised with an acute MI	Ongoing	No; population not relevant for the decision problem
EMPRISE	NCT03363464 (133) EUPAS20677 (134)	Real world comparative effectiveness, safety, healthcare resource utilisation and costs of empagliflozin *versus* DPP-4 inhibitors in T2DM in routine clinical care	Ongoing	No; population not relevant for the decision problem

Abbreviations: CKD, chronic kidney disease; CV, cardiovascular; HF, heart failure; HFrEF, heart failure with reduced ejection fraction; HFpEF, heart failure with preserved ejection fraction; MI, myocardial infarction; PROs, patient-reported outcomes; T2DM, type 2 diabetes mellitus

Of the studies listed in Table 9, the EMPEROR-Reduced trial provides the main evidence base for clinical efficacy and safety of empagliflozin in the population of HF patients with reduced left ventricular ejection fraction (LVEF ≤ 40%).

EMPEROR-Reduced (NCT03057977) was an international phase III trial from the EMPOWER programme that investigated the effect of empagliflozin versus placebo in addition to guideline-recommended therapy on the combined risk of CV death and HHF in approximately 3,730 patients with chronic HF and reduced LVEF (LVEF ≤ 40%), with or without diabetes. It also evaluated the effects of empagliflozin on recurrent hospitalisation events, renal function, CV death, all-cause mortality, and change in KCCQ clinical summary score (124). The rational for the design of EMPEROR-Reduced was that the data from EMPA-REG OUTCOME was not sufficient to demonstrate efficacy of empagliflozin in patients with chronic HF, especially those at increased risk of an outcome event. EMPEROR-Reduced trial was therefore enriched for patients with a markedly reduced ejection fraction and increased levels of natriuretic peptides as specified in the inclusion criteria

Table 11). The trial enrolled patients from nine UK sites, increasing its relevance to the NHS clinical practice. External validity of the trial is strengthened by the protocol requirement for patients to receive background HF treatment as per local guidelines. The generalisability of the trial results to the NHS clinical practice is discussed further in sections B.2.7, B.2.12, and B.3.2. Its outcomes provide the key clinical and QoL inputs for the economic model of empagliflozin in HFrEF. Design and methodology of EMPEROR-Reduced are described in section Table 10.

B.2.1.2 Non-randomised clinical effectiveness studies

Evidence from PULSE, a retrospective observational study of the burden of HF, including HFrEF, in England, was used to characterise patients seen in the NHS clinical practice and validate the long-term outcome predictions of the HFrEF cost-utility model for patients treated with the SoC against the real world outcomes. Patients with a diagnosis of HF recorded in the UK Clinical Practice Research Datalink (CPRD) or Hospital Episode Statistics (HES) database between 1 January 2015 and 31 December 2019 were eligible for inclusion in the PULSE study (43). Based on the availability of evidence of EF classification in CPRD records, the cohort was split into rEF, pEF and “unknown EF” subpopulations. The study objectives were to determine the incidence and prevalence of HF and HFrEF in England, estimate rates of HF- and HFrEF-related hospitalisation, CV and all-cause mortality, and evaluate resource utilisation over the study period. The outcomes of the PULSE study are therefore relevant to the decision problem considered in the cost-utility analysis of empagliflozin (Section B.3.2).

B.2.2 List of relevant clinical effectiveness evidence

The clinical evidence on empagliflozin as an addition to SoC in the treatment of chronic HFrEF consists of one phase III trial, EMPEROR-Reduced (Table 10). This pivotal trial was the main source of clinical efficacy evidence in the cost-utility model described in section B.3.

Table 10. Clinical effectiveness evidence: EMPEROR-Reduced trial

Study	EMPEROR-Reduced (NCT03057977) (124)
Primary sources	Packer et al 2020 (109); Butler et al 2021 (120)
Additional sources	EMPEROR-Reduced CSR (135)
Study design	• Phase III, randomised, double-blind, placebo-controlled trial with parallel assignment • The trial was event-driven and all randomised patients remained in the trial until the defined number of adjudicated primary endpoint events had been reached
Population	Adults with chronic HF NYHA class II-IV and reduced EF (LVEF ≤ 40%) who have been diagnosed at least 3 months before screening, with or without DM • N=3730 • Age ≥ 18 years • Baseline natriuretic peptide levels higher than a pre-specified concentration depending on the baseline EF (see Section B.2.3.1.2) • Appropriate dose of medical therapy for HF consistent with local and international guidelines, stable for at least one week prior to screening and during the screening period until randomisation (between 4 to 28 days)
Intervention(s)	Empagliflozin PO 10 mg once daily in addition to SoC (which could include treatment with an ACEi, ARB, mineralocorticoid receptor antagonist, beta-blocker and/or sacubitril valsartan)
Comparator(s)	Placebo plus appropriate background medical therapy for HF
Does trial support application for marketing authorisation?	Yes
Is trial used in the economic model?	Yes

Study	EMPEROR-Reduced (NCT03057977) (124)
Reported outcomes specified in the decision problem	The outcomes relevant for the decision problem include: • Time to first adjudicated CV death or adjudicated hospitalisation for HF • Occurrence of adjudicated hospitalisation for HF • Decline in renal function • Time to first occurrence of chronic dialysis, renal transplant or sustained reduction of eGFR • Time to first adjudicated hospitalisation for HF • Cardiovascular mortality • All-cause mortality • Occurrence of all-cause hospitalisation • Adverse effects of treatment • Patient-reported outcome measured by KCCQ • Health-related quality of life measured by EQ-5D-5L
All other reported outcomes	• New onset of atrial fibrillation • Adjudicated MI (fatal or non-fatal) • Adjudicated stroke (fatal or non-fatal) • Adjudicated TIA • Incidence of acute renal failure • Change from baseline in systolic blood pressure over time • Change from baseline in diastolic blood pressure • Change from baseline in pulse rate over time • Change from baseline in HbA1c over time

Abbreviations: ACEi, angiotensin-converting enzyme inhibitor; ARB, angiotensin II receptor blockers; CSR, clinical study report; CV, cardiovascular; DM, diabetes mellitus; EF, ejection fraction; EQ-5D, EuroQol five dimensions questionnaire; HbA1c, glycated haemoglobin; HF, heart failure; eGFR, estimated glomerular filtration rate; KCCQ, Kansas City Cardiomyopathy Questionnaire; LVEF, left ventricular ejection fraction; NYHA, New York Heart Association classification; MI, myocardial infarction; PO, per os; SoC, standard of care; TIA, transient ischaemic attack.

B.2.3 Summary of methodology of the relevant clinical effectiveness evidence

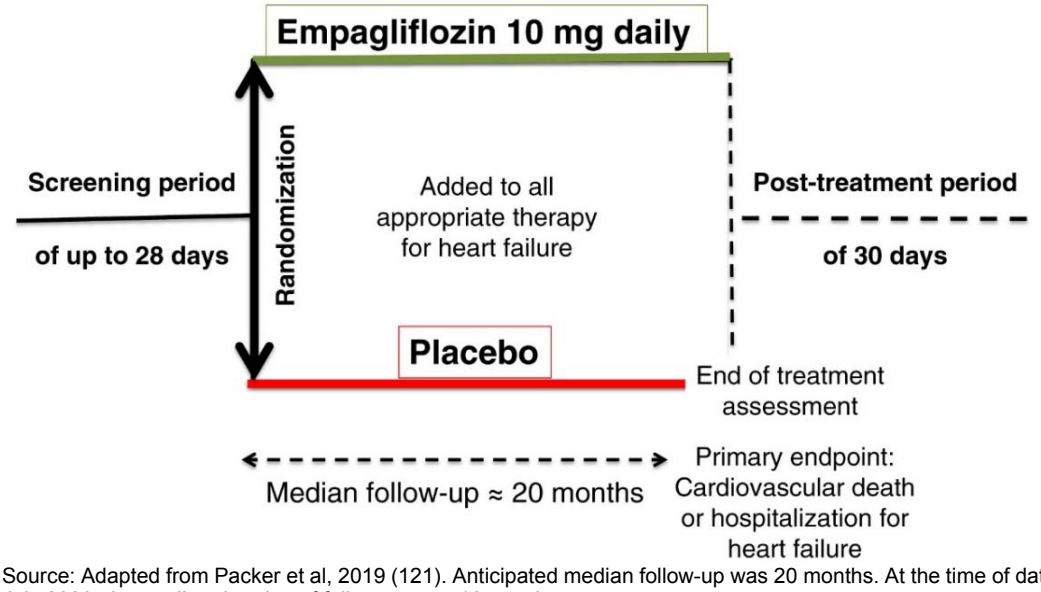
B.2.3.1 Summary of methodology of the EMPEROR-Reduced trial (NCT03057977)

EMPEROR-Reduced was an international phase III study designed to evaluate the long-term efficacy and safety of empagliflozin *versus* placebo in addition to guideline-directed medical therapy in patients with symptomatic HFrEF (LVEF ≤40%) (121). The trial had a double-blind, placebo-controlled, randomised design with parallel assignment of participants in 1:1 ratio to one of the two treatment arms:

- Empagliflozin, 10 mg PO once daily in addition to SoC (a guideline-directed medical therapy with ACEi/ARB or ARNi, beta-blockers and mineralocorticoid receptor antagonists), *or*
- Placebo PO once daily in addition to the SoC.

The trial design is illustrated in Figure 5. Following a screening period lasting 4–28 days, patients who fulfilled all eligibility criteria were randomised to receive placebo or empagliflozin daily in addition to their usual therapy for HF.

Figure 5. Design of EMPEROR-Reduced trial



Source: Adapted from Packer et al, 2019 (121). Anticipated median follow-up was 20 months. At the time of database lock on 14 July 2020, the median duration of follow-up was 16 months.

Randomisation was performed using a permuted block design with a computer pseudo-random number generator and was stratified by:

- geographical region (North America, Latin America, Europe, Asia or “Other”),
- history of diabetes (diabetes, pre-diabetes and no diabetes), and
- estimated glomerular filtration rate [eGFR by the Chronic Kidney Disease - Epidemiology Collaboration Equation (CKD-EPI) equation] at screening <60 or ≥ 60mL/min/1.73 m².

Following randomisation, all appropriate treatments for HF or other medical conditions were initiated and individualised at the discretion of each subject’s physician. Patients were evaluated periodically at pre-specified study visits.

The primary objective of the EMPEROR-Reduced was to compare the time to first event of adjudicated CV death or adjudicated HHF among patients taking empagliflozin relative to those taking placebo in addition to their standard cardio-renal-metabolic therapy. The trial also evaluated the effects of empagliflozin on recurrent HHF, renal function, cardiovascular death, all-cause mortality, and QoL.

EMPEROR-reduced was an event-driven trial and all randomised patients remained in the study until the defined number of adjudicated primary endpoint events were reached. As such, EMPEROR-Reduced was appropriately designed to determine if the addition of empagliflozin can improve outcomes of HFrEF relative to current approaches that have established benefits in the treatment of chronic HF with reduced ejection fraction. Aspects of the trial methodology are described in more detail below in accordance with the CONSORT statement (136).

B.2.3.1.1 Changes to trial design

The description of EMPEROR-Reduced methodology outlined in this submission is based on the revised study protocol number c09098452-04 which was issued on 20 November 2019 and incorporates Global Amendment 3.0.

B.2.3.1.2 Eligibility criteria for study participants

Adult patients with HFrEF (LVEF ≤ 40%) diagnosed at least 3 months before screening and in the functional NYHA class II-IV were eligible for enrolment in EMPEROR-

Reduced. The intent of the trial was to recruit HFrEF patients whose expected event rate for the combined risk of CV death and HHF was at least 15% per year. The trial protocol therefore required that baseline levels of N-terminal prohormone B-type natriuretic peptide (NT-pro-BNP) exceed pre-defined levels which varied depending on the EF, as described in

Table 11.

Table 11. Inclusion and exclusion criteria of the EMPEROR-Reduced trial

Inclusion criteria	• Males and females aged ≥ 18 years of age • Patients with chronic HF diagnosed for at least 3 months before Visit 1 (screening), and currently in HF NYHA class II-IV • Chronic HF with reduced EF defined as LVEF ≤ 40% per local reading • In addition to LVEF≤40%, patient must have at least one of the following (analysed at the Central Laboratory at screening): o If EF is ≤30%, elevated [NT-pro-BNP] ≥600 pg/mL in patients without AF OR ≥1200 pg/mL in patients with AF o If EF is ≥31 to ≤35%, elevated [NT-pro-BNP] ≥1000 pg/mL in patients without AF OR ≥2000 pg/mL in patients with AF o If the EF is ≥36 to ≤40%, elevated [NT-pro-BNP] ≥2500 pg/mL in patients without AF OR ≥5000 pg/mL in patients with AF o For EF ≤ 40% and documented HHF within 12 months prior to screening, elevated [NT-pro-BNP] ≥600 pg/mL in patients without AF and ≥1200 pg/mL in patients with AF • Appropriate dose of medical therapy for HF (such as ACEi, ARB, beta-blocker, oral diuretics, MRA, ARNI, ivabradine) consistent with prevailing local and international CV guidelines, stable for at least 1 week prior to Visit 1 (screening visit) and during screening period until Visit 2 (randomisation visit). The dose of diuretics must be stable for only one week prior to Visit 2 to control symptoms. • Appropriate use of medical devices such as cardioverter-defibrillator (ICD) or a cardiac resynchronisation therapy (CRT) consistent with prevailing local or international CV guidelines. • Body-Mass Index (BMI) < 45 kg/m^2 at Visit 1 (screening).
Exclusion criteria	• CV diseases or treatments that increase the unpredictability of or change the patients' clinical course, independent of HF o Myocardial infarction (increase in cardiac enzymes in combination with symptoms of ishemia or new ishemic ECG changes), CABG or other major CV surgery, stroke or transient ischaemic attack in past 90 days o Heart transplant recipient or listed for heart transplant. Currently implanted left ventricular assist device (LVAD)

	o Cardiomyopathy based on infiltrative diseases (amyloidosis), accumulation diseases (haemochromatosis, Fabry disease), muscular dystrophies, cardiomyopathy with reversible causes (e.g. stress cardiomyopathy), hypertrophic obstructive cardiomyopathy or known pericardial constriction o Diagnosis of peripartum cardiomyopathy or cardiomyopathy induced by chemotherapy within 12 months o Any severe (obstructive or regurgitant) valvular heart disease expected to lead to surgery during the trial period o Acute decompensated heart failure requiring intravenous diuretics, vasodilators, inotropic agents or mechanical support within 1 week of screening and during the screening period prior to randomisation o ICD or cardiac resynchronisation therapy within 3 months prior to screening or if there is an intent to implant either device for 3 months following screening • Untreated or undertreated CV conditions that might influence the course of HF or tolerability of the study medications o Atrial fibrillation or atrial flutter with a resting heart rate >110 bpm, documented by ECG at screening o Untreated ventricular arrhythmia with syncope in a patient without an ICD within 3 months prior to screening o Symptomatic bradycardia or second or third-degree heart block without a pacemaker after adjustment of beta-blocker therapy, if appropriate o Systolic blood pressure ≥180mmHg at randomisation. If systolic blood pressure is 151–179mmHg, the patient should be receiving ≥3 anti-hypertensive drugs o Symptomatic hypotension and/or a systolic blood pressure <100mmHg at screening or at randomisation • Significant comorbidities that might influence the clinical course independent of HF o Chronic PD requiring home oxygen, oral corticosteroid therapy or hospitalisation for exacerbation within 12 months, significant chronic PD or primary pulmonary arterial hypertension o Acute or chronic liver disease, defined by serum levels of transaminase or alkaline phosphatase more than three times the upper limit of normal at screening o Impaired renal function, defined as eGFR <20mL/min/1.73m^2 (CKD-EPI) or requiring dialysis at the time of screening o Haemoglobin <9 g/dL at screening o Major surgery performed within 90 days prior to screening or major scheduled elective surgery (e.g. hip replacement) within 90 days after screening

	o GI surgery or GI disorder that could interfere with medication absorption o Any documented active or suspected malignancy or history of malignancy within 2 years prior to screening, except appropriately treated basal cell carcinoma of the skin, *in situ* carcinoma of uterine cervix or low risk prostate cancer o Presence of any other disease with a life expectancy of <1 year (in the opinion of the investigator) • Any condition that might jeopardise patient safety, limit the patient's participation in the trial or undermine the interpretation of trial data o Current use or prior use of a SGLT2i or combined inhibitor of SGLT1 and SGLT2 within 12 weeks prior to screening or randomisation o Discontinuation of a SGLT2i or combined inhibitor of SGLT1 and SGLT2 for the purposes of study enrolment is not permitted o Known allergy or hypersensitivity to any SGLT2is o History of ketoacidosis o Patients who must or wish to continue the intake of restricted medications or any drug considered likely to interfere with the safe conduct of the trial o Currently enrolled in another investigational device or drug study or are less than 30 days since the completion of a trial of another investigational device or drug. Any patient receiving any investigational treatment other than the study medications for this trial o Chronic alcohol or drug abuse or any condition that, in the investigator's opinion, will make the patient unlikely to fulfil the trial requirements or complete the trial o Women who are pregnant or are nursing or who plan to become pregnant while in the trial o Any other clinical condition that would jeopardise patient safety while participating in this trial or may prevent the subject from adhering to the trial protocol

Abbreviations: AF, atrial fibrillation; bpm, beats per minute; CABG, coronary artery bypass grafting; CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration equation; CV, cardiovascular; ECG electrocardiogram; EF, ejection fraction; eGFR, estimated glomerular filtration rate;GI, gastrointestinal; ICD, implantable cardioverter-defibrillator; HF, heart failure; LVEF, left ventricular ejection fraction; NYHA, New York Heart Association classification; NT-pro-BNP, N-terminal prohormone B-type natriuretic peptide; PD, pulmonary disease; SGLT, sodium-glucose co-transporter; SGLT2i, sodium-glucose co-transporter 2 inhibitor.

B.2.3.1.3 Study locations

Patient enrolment (N=3730) started on 6 March 2017 in university hospitals, specialist cardiovascular clinics and clinical research centres across 520 locations in 20 countries (US, Argentina, Australia, Belgium, Brazil, Canada, China, Czech Republic, France, Germany, Hungary, India, Italy, Japan, Republic of Korea, Mexico,

Netherlands, Poland, Spain, and the UK). The study completion date was 28 May 2020.

B.2.3.1.4 Trial drugs and concomitant medications

Study interventions are summarised in Table 12. At any time during the treatment period the investigator could adjust and optimise HF background therapy according to local and international guidelines. If additional therapy was necessary during the treatment period, it could be given at the discretion of the investigator. Disallowed concomitant medications included any SGLT2 inhibitors or combined SGLT-1 and 2 inhibitors, except the blinded trial medication.

Table 12. EMPEROR-Reduced trial drugs

Drug	Dose	Frequency of administration	Route of administration	Duration
Empagliflozin, film coated tablet	10 mg	Once daily	Oral	Until the necessary number of events were observed to evaluate efficacy for the primary composite endpoint
Placebo matching empagliflozin, film coated tablet	-			

B.2.3.1.5 Pre-specified primary and secondary outcomes of EMPEROR-Reduced

The endpoints relevant for the decision problem are summarised in Table 13. The definitions of adjudicated CV endpoints are summarised in Table 14.

Table 13. Pre-specified primary and secondary outcomes

Primary endpoint	Definition	NICE scope/ economic model?
Combined risk of CV death or hospitalisation for HF (HHF)	Time to first event analysis of the combined risk of adjudicated CV death or adjudicated HHF	Per NICE scope; not included in the economic model as a composite outcome

Key secondary endpoints	Definition	NICE scope/economic model?
Total HHF (first and recurrent)	Occurrence of adjudicated HHF (first and recurrent)	Per NICE scope; included in the economic model
Rate of renal function decline	eGFR (CKD-EPI)cr slope of change from baseline	Per NICE scope; not included in the economic model
Other secondary endpoints	**Definition**	**NICE scope/economic model?**
Risk of composite renal endpoint (chronic dialysis, renal transplant or renal insufficiency)	Time to first event in the composite renal endpoint: occurrence of chronic dialysis† or renal transplant or sustained¶ reduction§ in eGFR (CKD-EPI)cr from baseline of ≥ 40% or • sustained eGFR (CKD-EPI)cr <15 mL/min/1.73 m^2 for patients with baseline eGFR ≥30 mL/min/1.73 m^2 or • sustained eGFR (CKD-EPI)cr <10 mL/min/1.73 m^2 for patients with baseline eGFR <30 mL/min/1.73 m^2	Per NICE scope; included in the economic model
Risk of first HHF	Time to first adjudicated HHF	Per NICE scope; not included in the economic model
Risk of CV death	Time to adjudicated CV death	Per NICE scope; included in the economic model
Risk of death	Time to all-cause mortality	Per NICE scope; included in the economic model
Risk of diabetes mellitus	Time to onset of DM defined as HbA1c ≥6.5% or as diagnosed by the Investigator in patients with pre-DM (defined as no history of DM and no HbA1c ≥6.5% before treatment, and a pre-treatment HbA1c value of ≥ 5.7% to <6.5%)	Not in scope; not cluded in the economic analysis
Change in KCCQ clinical summary score	Change from baseline in the KCCQ clinical summary score (HF symptoms and physical limitations domains) at week 52	Per NICE scope; included in the economic model

Risk of all-cause hospitalisation	Occurrence of all-cause hospitalisation (first and recurrent)	Per NICE scope; not included in the economic model
Further endpoints	**Definition**	**NICE scope/economic model?**
Risk of atrial fibrillation	New onset of atrial fibrillation	Not in scope; not included in the economic model
Risk of myocardial infarction	Adjudicated myocardial infarction (fatal or non-fatal)	Not in scope; not included in the economic model
Risk of stroke	Adjudicated stroke (fatal or non-fatal)	Not in scope; not included in the economic model
Safety	Adverse events, adverse events of special interest, and specific adverse events	Per NICE scope; included in the economic model

Abbreviations: CV, cardiovascular; CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration equation; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; HbA1c, haemoglobin A1c; HF, heart failure; HHF, hospitalisation for heart failure; KCCQ, Kansas City Cardiomyopathy Questionnaire

†Chronic dialysis was defined as dialysis with a frequency of twice per week or more for at least 90 days

¶Sustained was determined by two or more consecutive post-baseline central laboratory measurements separated by at least 30 days (the first to last of the consecutive eGFR values)

§Reduction in eGFR (CKD-EPI)cr was defined as reduction in eGFR from baseline of ≥40%, eGFR <15 mL/min/1.73m^2 for patients with baseline eGFR ≥30 mL/min/1.73m^2, or eGFR <10 mL/min/1.73m^2 for patients with baseline eGFR <30mL/min/1.73m^2

Table 14. Definitions of adjudicated endpoints

Endpoint	Definition¶
HHF	HHF endpoint must meet the following criteria: • Adjudicated primary diagnosis is admission to hospital for HF • Length of stay in hospital extends for ≥12 hours (emergency room visit for ≥12 hours with IV therapy is considered equivalent to admission to hospital) • The patient exhibits documented new or worsening symptoms due to HF on presentation, including at least one of the following: o Dyspnoea (dyspnoea with exertion, dyspnoea at rest, orthopnoea, paroxysmal nocturnal dyspnoea) o Decreased exercise tolerance o Fatigue o Other symptoms of worsened end-organ perfusion (dizziness, confusion, or volume overload such as weight gain or lower extremity swelling)

Endpoint	Definition[¶]
	• Objective evidence of new or worsening HF consisting of at least two physical examination findings or one physical examination finding and at least one laboratory criterion, including: o Physical examination findings considered to be due to HF: - Peripheral oedema - Increasing abdominal distension or ascites - Pulmonary rales/crackles/crepitations - Increased jugular venous pressure and/or hepatojugular reflux - S3 gallop - Clinically significant rapid weight gain related to fluid retention o Laboratory evidence of new or worsening HF, if obtained within 24 hours of presentation, including: - Increased BNP/NT pro-BNP concentrations consistent with decompensation of HF - Radiological evidence of pulmonary congestion - Non-invasive evidence of clinically significant left- or right- sided ventricular filling pressure or low cardiac output, *or* - Invasive diagnostic evidence with right heart catheterisation showing a pulmonary capillary wedge pressure ≥18 mmHg, central venous pressure ≥12 mmHg, or a cardiac index <2.2 L/min/L^2 • The patient receives initiation or intensification of treatment for HF, including at least one of the following: o Augmentation in oral diuretic therapy o IV diuretic or vasoactive agent (e.g. inotrope, vasopressor, or vasodilator) o Mechanical or surgical intervention (circulatory support with intra-aortic balloon pump, ventricular assist device, extracorporeal membrane oxygenation, total artificial heart or fluid removal with ultrafiltration, hemofiltration, dialysis)
CV death	CV death includes the following categories: • Death due to MI, a procedure to treat MI or elective coronary procedure to treat myocardial ischemia • Death due to clinically worsening signs and symptoms of HF including cardiogenic shock and pulmonary edema • Death due to stroke, CV procedures, CV haemorrhage or other CV causes (e.g. pulmonary embolism or peripheral arterial disease) • Sudden cardiac death, including: o Death witnessed and occurring without new or worsening symptoms

Endpoint	Definition[¶]
	o Death witnessed within 60 minutes of the onset of new or worsening cardiac symptoms o Death witnessed and attributed to an identified arrhythmia or unwitnessed but found on implantable cardioverter-defibrillator review o Death after unsuccessful resuscitation from cardiac arrest or successfully resuscitated from cardiac arrest without identification of a specific cardiac or non-cardiac aetiology o Unwitnessed death in a subject seen alive and clinically stable ≤ 72 hours prior to being found dead without any evidence supporting a specific non-cardiovascular cause of death

Abbreviations: CV, cardiovascular; HF, heart failure; HHF, hospitalisation for heart failure; BNP, B-type natriuretic peptide; NT pro-BNP, N-terminal pro-B- type natriuretic peptide; MI, myocardial infarction

[¶]All CV endpoint definitions were modifications of the guideline recommendations by Hicks et al 2014 (137).

B.2.3.2 Demographics and baseline characteristics

Patients in the empagliflozin and the placebo group were well balanced with respect to demographic and clinical characteristics at baseline (Table 15). Overall, a quarter of patients had HF NYHA class III-IV, 73% had LVEF of 30% or less, 79% had a NT pro-BNP level of at least 1000 pg/ml, 48% had an estimated GFR of less than 60 ml per minute per 1.73 m^2 and nearly 20% were receiving an angiotensin receptor-neprilysin inhibitor.

Table 15. Demographic and baseline characteristics (mean) of randomised participants in EMPEROR-Reduced trial

Baseline characteristic*	Empagliflozin 10 mg	Placebo
Number of subjects	1863	1867
Age (years), mean (SD)	67.2±10.8	66.5±11.2
Female sex, No (%)	437 (23.5)	456 (24.4)
Race, No (%)[†]		
White	1325 (71.1)	1304 (69.8)
Black	123 (6.6)	134 (7.2)
Asian	337 (18.1)	335 (17.9)
Other or missing	78 (4.2)	94 (5.0)

Baseline characteristic*	Empagliflozin 10 mg	Placebo
Region, No (%)		
North America	212 (11.4)	213 (11.4)
Latin America	641 (34.4)	645 (34.5)
Europe	676 (36.3)	677 (36.3)
Asia	248 (13.3)	245 (13.1)
Other	86 (4.6)	87 (4.7)
NYHA functional class, No (%)		
II	1399 (75.1)	1401 (75.0)
III	455 (24.4)	455 (24.4)
IV	9 (0.5)	11 (0.6)
Body-mass index‡ (kg/m^2), mean (SD)	28.0±5.5	27.8±5.3
Heart rate (beats/min), mean (SD)	71.0±11.7	71.5±11.8
SBP (mm Hg), mean (SD)	122.6±15.9	121.4+15.4
DBP (mm Hg), mean (SD)	74.0 (11.0)	73.7 (10.6)
Left ventricular ejection fraction		
Mean (SD)	27.7±6.0	27.2±6.1
Value of ≤ 30%, No (%)	1337 (71.8)	1392 (74.6)
NT pro-BNP		
Median (IQR) (pg/ml)	1887 (1077-3429)	1926 (1153-3525)
Value of ≥1000 pg/ml, No/total No (%)	1463/1862 (78.6)	1488/1866 (79.7)
Cause of heart failure, No (%)		
Ischaemic	983 (52.8)	946 (50.7)
Nonischaemic	880 (47.2)	921 (49.3)

Baseline characteristic*	Empagliflozin 10 mg	Placebo
Cardiovascular history, No (%)		
Hospitalisation for HF in ≤12 months	577 (31.0)	574 (30.7)
Atrial fibrillation	664 (35.6)	705 (37.8)
Diabetes mellitus	927 (49.8)	929 (49.8)
Hypertension	1349 (72.4)	1349 (72.3)
Estimated glomerular filtration rate		
Mean (SD) (ml/min/1.73 m^2)	61.8 ± 21.7	62.2 ± 21.5
Value of <60 ml/min/1.73 m^2, No/total No (%)	893/1862 (48.0)	906/1866 (48.6)
UACR (mg/ml), N (%)		
Normal (<30)	1038 (55.7)	1040 (55.7)
Microalbuminuria (30 to ≤300)	608 (32.6)	628 (33.6)
Macroalbuminuria (>300)	207 (11.1)	189 (10.1)
Heart failure medication, No (%)		
Renin–angiotensin inhibitor§		
Without neprilysin inhibitor	1314 (70.5)	1286 (68.9)
With neprilysin inhibitor	340 (18.3)	387 (20.7)
Mineralocorticoid receptor antagonist	1306 (70.1)	1355 (72.6)
Beta-blocker	1765 (94.7)	1768 (94.7)
Device therapy, No (%)		
Implantable cardioverter-defibrillator¶	578 (31.0)	593 (31.8)
Cardiac resynchronisation therapy‖	220 (11.8)	222 (11.9)

Baseline characteristic*	Empagliflozin 10 mg	Placebo
Diabetes status		
Without diabetes, N (%)	936 (50.2)	938 (50.2)
Without diabetes or pre-diabetes, N (%)	304 (16.3)	302 (16.2)
With pre-diabetes, N (%)	632 (33.9)	636 (34.1)
With diabetes, N (%)	927 (49.8)	929 (49.8)
T2DM, N (%)	927 (49.8)	929 (49.8)
T1DM, N (%)	0	0

Abbreviations: DBP, diastolic blood pressure; HF, heart failure; IQR, interquartile range; No, number; NT pro-BNP, N-terminal prohormone of brain natriuretic peptide; NYHA, New York Heart Association; SBP, systolic blood pressure; SD, standard deviation; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus; UACR, urine albumin-to-creatinine ratio.
* Plus-minus values are means ± SD. Percentages may not total 100 because of rounding.
† Race was reported by the patients. Those who identified with more than one race or with no race were classified as "other".
‡ The body-mass index is the weight in kilograms divided by the square of the height in meters.
§ Inhibitors of the renin–angiotensin system include angiotensin converting enzyme inhibitors and angiotensin receptor blockers.
¶ This category includes all the patients with an implantable cardioverter-defibrillator regardless of the presence or absence of cardiac resynchronisation therapy.
‖ This category includes all the patients who were receiving cardiac resynchronisation therapy regardless of the presence or absence of a defibrillator.

B.2.4 Statistical analysis and definition of study groups in the relevant clinical effectiveness evidence

The statistical analysis methods and definitions of study groups used in the pivotal EMPEROR-Reduced trial are described in Table 16.

B.2.4.1 Statistical methods and analysis sets

Table 16. Summary of statistical analysis in the EMPEROR-Reduced RCT

Study name (number)	EMPEROR-Reduced (NCT03057977)
Research hypothesis	There is no difference between the efficacy of empagliflozin and efficacy of placebo in reducing the combined risk of CV death and HHF
Analysis sets	• **Screened Set (SCR):** All patients screened for the trial, with informed consent given and who completed at least one screening procedure at Visit 1 • **Randomised set (RS):** All randomised patients, whether treated or not • **Treated set (TS):** All patients who were dispensed study medication and were documented to have taken at least one dose of investigational treatment

Study name (number)	EMPEROR-Reduced (NCT03057977)
	• **Treated Set-Follow-up (TS-FU):** All patients in the TS for whom a follow-up visit was performed (i.e. values of planned assessments: KCCQ, EQ-5D, vital signs or lab data reported) between 23 and 45 days after last intake of study medication. The TS-FU did not include patients for whom no planned measurements were taken, which happened in case of telephone FU visits. Patients with intake of open label SGLT2 inhibitor between their EOT and FU visit were also excluded from the TS-FU set
Statistical analysis for primary endpoint	The primary and key secondary endpoints were tested in the following hierarchical order: • Time to first event of adjudicated CV death or adjudicated HHF • Occurrence of adjudicated HHF (first and recurrent) • eGFR (CKD-EPI)cr slope of change from baseline For each of these confirmatory endpoints, superiority of empagliflozin over placebo was evaluated with a two-sided test. The overall type I error rate for the trial was preserved at α = 0.05. Due to the amount of α spent on the interim analysis, the remaining two-sided α level for the final analysis was 0.0496. The primary analysis was a Cox PH regression with factors treatment, geographical region, diabetes status at baseline, age, gender, LVEF, and baseline eGFR (CKD-EPI)cr. Following the ITT principle, the primary analysis was based on RS using all data up to the end of the planned treatment period (i.e. excluding events and time at risk after the protocol-specified treatment discontinuation for patients who completed the treatment period but including the data after end of treatment for patients not completing the treatment phase as planned). Patients without a specific endpoint event were censored at the last date the patient was known to be event free or at the end of the planned treatment period, whichever was earlier. When violation of the PH assumption was observed, groups of patients for which the proportionality assumption held were identified, and a stratified Cox regression was performed.
Statistical analysis for key secondary endpoints	• **Occurrence of adjudicated HHF (first and recurrent)** was analysed by a joint frailty model that accounted for the dependence between recurrent HHF and CV death. The primary analysis included all data until completion of the planned treatment phase, including the data after end of treatment for patients not completing the treatment phase as planned. The model included the same covariates used for the analysis of the primary endpoint. The joint model provided two distinct HRs: o HR_{HHF} associated with the effect of treatment on the recurrent event rate of HHF o HR_{CVD}, the hazard ratio for CV death. • **Slope in change from baseline of eGFR (CKD-EPI)cr** was analysed by a random coefficient model allowing for random intercept and random slope per patient, with the same factors used for the primary endpoint and the additional factors time, treatment-by-time and baseline eGFR (CKD-EPI)cr-by-time interaction as linear

Study name (number)	EMPEROR-Reduced (NCT03057977)
	covariates. The model included all on-treatment change from baseline. This endpoint was tested with a two-sided α of 0.001.
Statistical analysis of exploratory endpoints	• **Time to event endpoints:** as analysis of primary endpoint • **Recurrent event endpoints:** as analysis of the first key secondary endpoint • **Continuous endpoints:** mixed model repeated measure analysis (MMRM) • **Categorial endpoints:** descriptive
Sample size & power calculation	Sample size calculation was based on the number of events needed to detect a 20% difference in risk of a primary endpoint event with 90% power for a two-sided test with α=0.05. Achieving that treatment effect size required 841 primary endpoint events. Assuming a ≈15% event rate per year in the placebo arm, a recruitment period of 18 months and a follow-up period of 20 months, 2,850 patients needed to be randomised to receive empagliflozin or placebo in 1:1 manner.
Data management, patient withdrawals	• **Handling of drop-outs or missing data:** o For patients without primary event and lost to follow-up before trial completion, the treatment specific incidence rates for empagliflozin and placebo for retrieved drop-outs were used to impute the primary events in a multiple imputations framework. The primary model was applied to the imputed datasets. o There was no imputation of data for safety analyses. o For endpoints of KCCQ scores in case of patients who die, a score of 0 was imputed at all subsequent scheduled visits where the score would have been assessed. o Missing covariates in multivariate Cox regression models and for recurrent event analyses were imputed using the overall population median of the corresponding variable for continuous covariates and the most frequent category for categorical covariates. No imputation was done for covariates included in treatment by subgroup interaction terms. • **Subjects could have been instructed to permanently discontinue study drug only after discussion with investigator if:** o eligibility criteria were violated o in the case of an AE o if the patient failed to comply with the protocol o if any restricted treatment was given during the trial

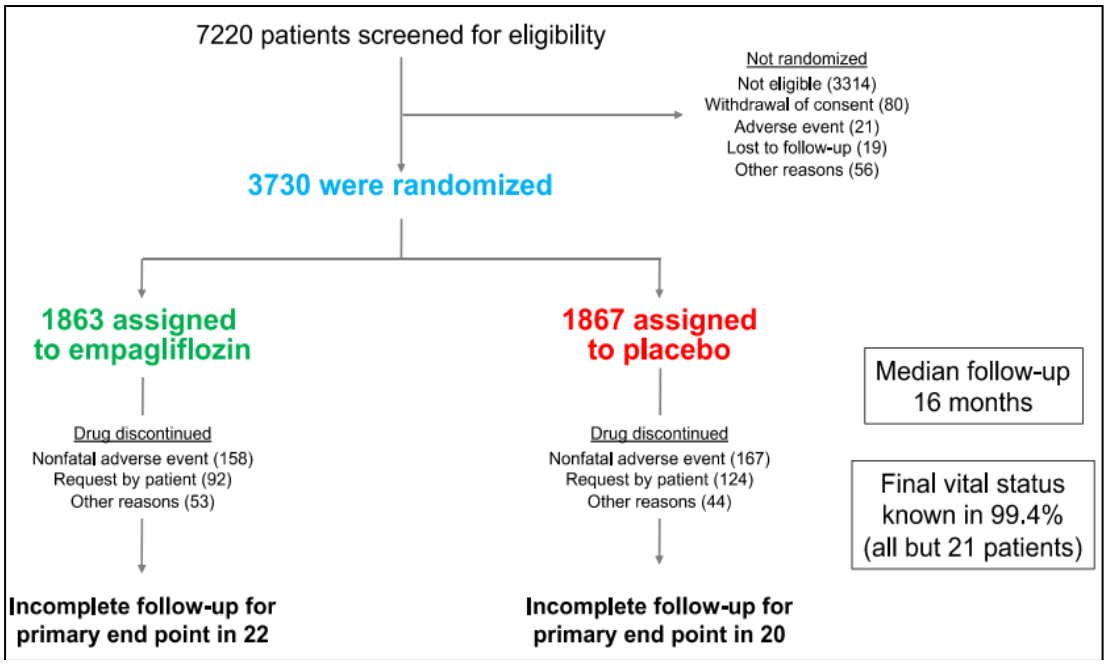
Source: Packer et al 2019 (121); EMPEROR-Reduced CSR (135)

Abbreviations: AE, adverse event; CI, confidence interval; CV, cardiovascular; CVD, cardiovascular death; (CKD-EPI)cr, Chronic Kidney Disease Epidemiology Collaboration equation based on creatinine measurement; CPRD, Clinical Practice Research Datalink; DM, diabetes mellitus; EF, ejection fraction; EOT, end of treatment; EQ-5D, EuroQol 5 dimensions instrument; FU, follow-up; HES, Hospital Episode Statistics; HHF, hospitalisation for heart failure; HF, heart failure; HFrEF, heart failure with reduced ejection fraction; HR, hazard ratio; eGFR, estimated glomerular filtration rate; ITT, intention to treat; KCCQ, Kansas City Cardiomyopathy Questionnaire; LVEF, left ventricular ejection fraction; MMRM, mixed model repeated measure analysis; PH, proportional hazards; RS, randomised set; SCR, screened set; SGLT2, sodium-glucose co-transporter 2; SoC, standard of care; TS, treated set; TS-FU, treated set with follow-up.

B.2.4.2 Participant flow in the relevant randomised controlled trials

Participant flow in EMPRERO-Reduced is shown in Figure 6.

Figure 6. CONSORT diagram of patient flow in each stage of EMPREOR-Reduced RCT



Note: Incomplete follow-up for the primary end point refers to incomplete information on either vital status or hospitalisation until the planned end of the treatment period for those patients who had not experienced an adjudicated primary outcome. The 21 patients with unknown vital status at the end of the trial included eleven on empagliflozin and ten on placebo. Three patients with missing vital status at the end of the trial experienced an adjudicated HHF and are not considered to have incomplete follow-up for the primary endpoint.

B.2.5 Quality assessment of the relevant clinical effectiveness evidence

A summary of the quality assessment of EMPEROR-Reduced, a parallel group RCT, is shown in

Table 17. The complete quality assessment is provided in Appendix D.

Table 17. Results of the quality assessment of EMPEROR-Reduced trial

	EMPEROR-Reduced (NCT03057977)
Was randomisation carried out appropriately?	Yes. Randomisation was performed by using a permuted block design with a computer pseudo-random number generator.
Was the concealment of treatment allocation adequate?	Yes. An Interactive Response Technology System (voice response or web response) was used to determine treatment assignment.
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes. Demographic and patient characteristics were well balanced between the two treatment groups at baseline, and randomisation was stratified by geographical region, diabetes status and eGFR at screening.
Were the care providers, participants and outcome assessors blind to the treatment allocation?	Yes. This was a double-blind study. An Endpoint Adjudication Committee evaluated all reported and potential clinical events in a manner blinded to the treatment assignment.
Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	No. Proportion of patients who discontinued study treatment was low and well balanced between the two treatment groups.
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No. All outcomes specified in the study protocol were reported in the clinical study report.
Did the analysis include an intention to treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes. Efficacy analysis were performed in the randomised set.

B.2.6 Clinical effectiveness results of the relevant trials: EMPEROR-Reduced

As described in sections that follow, the null hypotheses for the primary and the two key secondary endpoints of the EMPEROR-Reduced trial were rejected in a hierarchical testing procedure. Results of the trial demonstrate that empagliflozin is superior to placebo in improving HF outcomes in patients with symptomatically stable HFrEF (LVEF ≤40%) on baseline guideline-directed medical therapy, irrespective of diabetes status. Benefit is primarily driven by a reduction in HHF. Addition of empagliflozin to standard care is also associated with a slower rate of decline in the estimated glomerular filtration rate (eGFR) in comparison to placebo, resulting in a lower risk of serious renal outcomes (109).

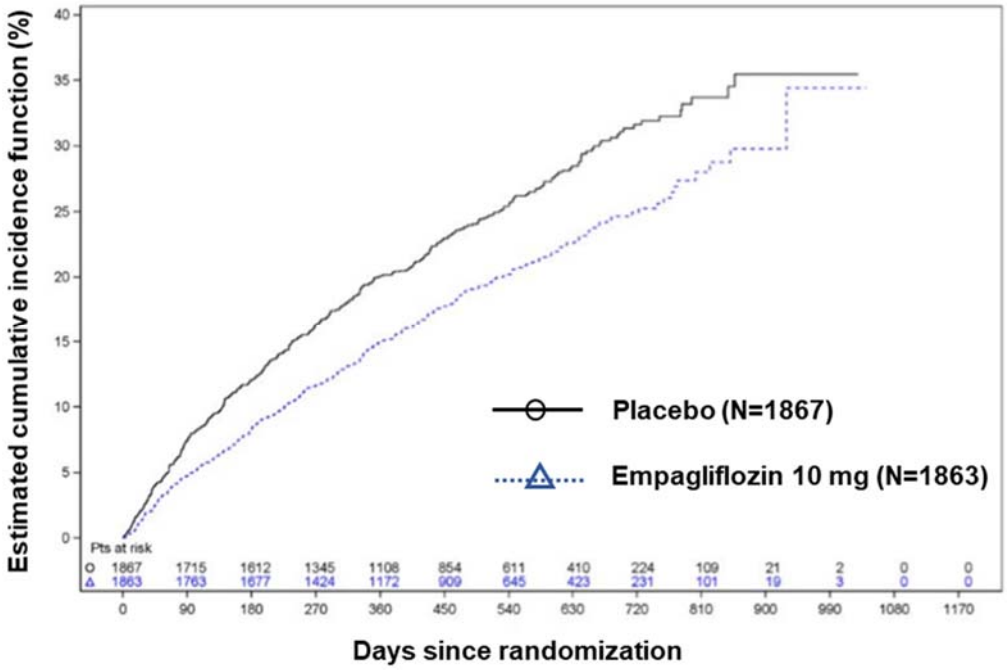
Importantly, during the trial period, benefits were observed in patients receiving standard care with any of the currently recommended drugs for HF, including sacubitril valsartan. Even patients with severe left ventricular dysfunction, with or without diabetes, appeared to benefit from addition of empagliflozin to their standard care.

Results of the pre-specified efficacy outcomes that are within the scope of the decision problem are described in sections B.2.6.1 to B.2.6.2 . Pre-specified subgroup analysis of the trial data used in the economic model is described in Appendix E.

B.2.6.1 Primary outcome: combined risk of cardiovascular death or HHF

Over a median follow-up of 16 months, the primary composite outcome of CV death or HHF occurred in a lower proportion of patients in the empagliflozin group (361 of 1863 patients, 19.4%) than in the placebo group (462 of 1867 patients, 24.7%). The separation of the estimated cumulative incidence of CV death or first HHF curves, considering non-CV death as a competing risk, started shortly after randomisation and was maintained throughout the trial period (Figure 7). Cox regression of data for all randomised patients adjusted for age, baseline eGFR (CKD-EPI)cr, region, gender, treatment, baseline diabetes status and LVEF, revealed that the risk of CV death or HHF was significantly reduced with empagliflozin compared with placebo (HR, 0.75; 95% CI, 0.65 to 0.86; p<0.0001).

Figure 7. Estimated cumulative incidence function for time to the first event of adjudicated CV death or HHF in all randomised patients (RS, randomised set)



Source: EMPEROR-Reduced CSR, Figure 11.1.1.1:1 (135)

During the trial period, lasting from the start of enrolment in April 2017 until the final follow-up data collection for the double-blind treatment period on 29 April 2020, the number of patients who needed to be treated with empagliflozin to prevent one primary event was 19 (95% CI, 13 to 37).

Several sensitivity analyses of the primary endpoint were performed to consider competing risks and to account for missing follow-up data in 42 patients who discontinued trial prematurely. The results were consistent with the results of the primary analysis, with HRs being numerically similar (Table 18).

Table 18. Sensitivity analyses for the primary endpoint: time to the first event of adjudicated CV death or HHF

Sensitivity analyses in RS	Hazard ratio (95% confidence interval)
Multiple imputation analysis addressing incomplete data for primary endpoint*, RS	0.75 (0.66-0.87)
Results unadjusted for covariates, RS	0.75 (0.66-0.86)
Sub-distribution hazard ratio adjusted for non-CV death as a competing risk in RS (Fine-Gray model)§	0.75 (0.66-0.86)

Source: Packer et al 2020 (109).
Abbreviations: CV, cardiovascular; HHF, hospitalisation for heart failure; RS, randomised set.
*Imputations were performed for 42 patients with incomplete data (20 placebo, 22 empagliflozin). Treatment specific incidence rates for empagliflozin and placebo for patients who discontinued study medication with available follow-up data were used to impute the primary events in a multiple imputations framework via sampling from an exponential distribution. One hundred imputations were performed and evaluated by the primary model. Log hazard ratios were summarised by Rubin's rules (138).
§Fine and Gray, 1999 (139).

B.2.6.2 Secondary outcomes

B.2.6.2.1 Hospitalisation for heart failure (first and recurrent)

The total number of hospitalisations for HF was lower in the empagliflozin group than in the placebo group with 388 events and 553 events, respectively. The mean cumulative incidence of HHF in the empagliflozin and placebo groups started to diverge shortly after randomisation and continued to segregate further over the course of the trial (

Figure 8). Primary analysis using joint frailty model with CV death as a competing risk demonstrated that the risk of recurrent HHF was significantly reduced with empagliflozin relative to placebo (HR, 0.70; 95%CI, 0.58 to 0.85, p<0.001). The hazard of recurrent HHF was positively correlated to that of CV death, as indicated by a frailty exponent greater than zero (data not shown).

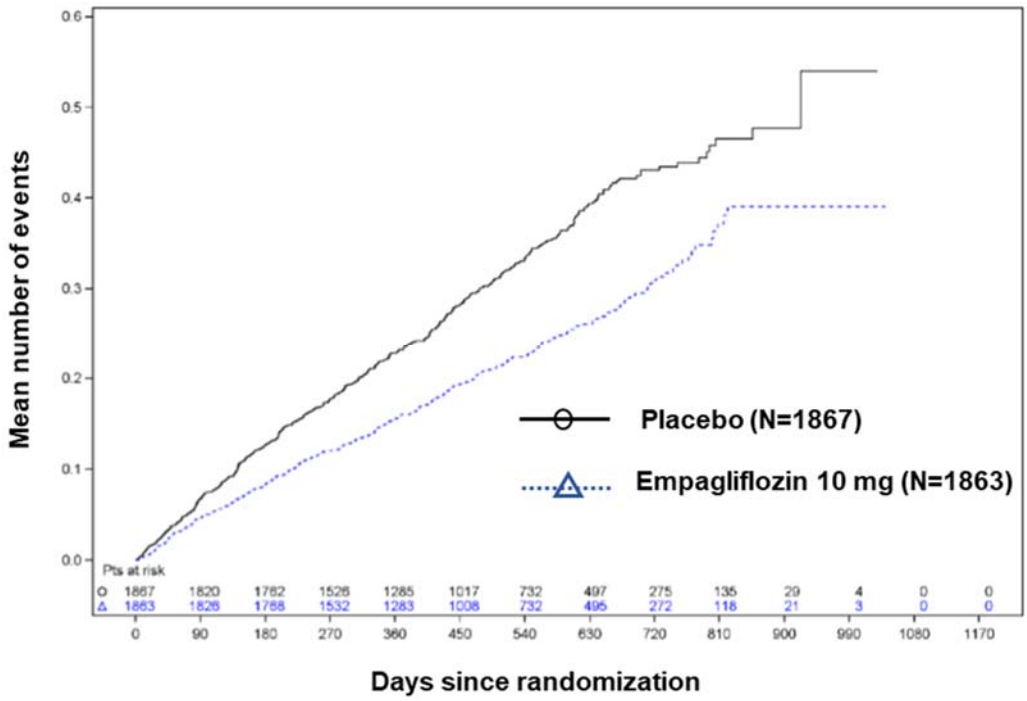
The results of the sensitivity analyses were consistent with the results of the primary analysis for the occurrence of adjudicated HHF (first and recurrent) (Table 19).

Table 19. Sensitivity analyses for the key secondary endpoint: total HHF

Sensitivity analyses	Hazard ratio (95% confidence interval)
Results unadjusted for covariates, RS¶	0.69 (0.57-0.84)
Using joint frailty model* with all-cause mortality instead of CV death as a competing risk, RS	0.70 (0.58-0.85)
Treated patients while on-treatment + 30 days, TS	0.69 (0.56-0.85)

Source: EMPEROR-Reduced CSR, Figure 11.1.2.1.2:1 (135)
Abbreviations: CV, cardiovascular; HF, heart failure; RS, randomised set; TS, treated set.
*Joint frailty model by Rogers et al. 2016 (140).
¶ Indicates post-hoc analysis.

Figure 8. Mean cumulative function for occurrence of adjudicated HHF (first and recurrent) in the RS



Source: EMPEROR-Reduced CSR, Figure 11.1.2.1.1:1 (135)

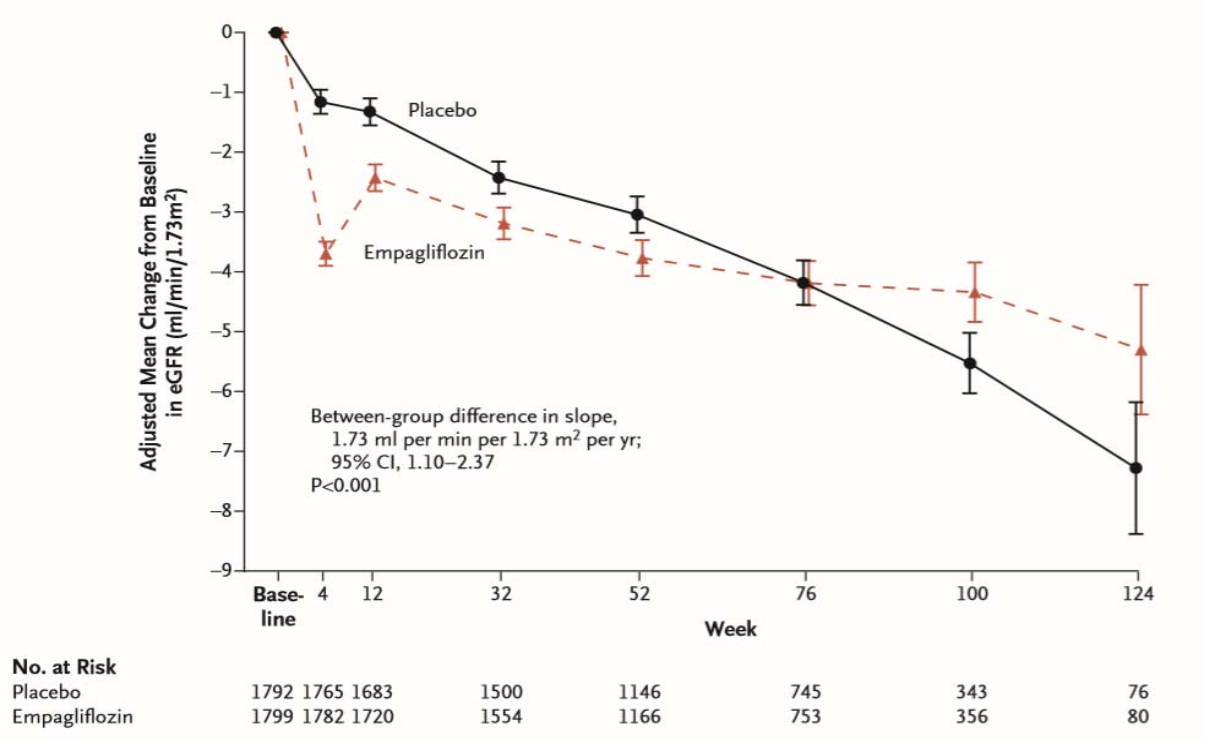
B.2.6.2.2 Deterioration of renal function

The other key secondary endpoint in the hierarchical testing procedure was mean slope of change in eGFR [mL/min/1.73 m^2] from baseline. Estimation of GFR was based on serum creatinine (cr) Chronic Kidney Disease Epidemiology Collaboration equation [(CKD-EPI)cr] (141).

The primary analysis included only "on-treatment" data from the treated set (TS) and measurements up to one day after the last intake of study medication. In the empagliflozin group, the estimated slope was -0.55 ± 0.23 mL/min/1.73m^2 per year. In the placebo group, eGFR declined more steeply over the duration of the treatment period, with an estimated slope of -2.278 ± 0.23 mL/min/1.73m^2 per year. The estimated between-group difference in mean slope was 1.73 mL/min/1.73m^2 per year (95% CI, 1.10 – 2.37; p<0.001) (

Figure 9). In the randomised set, the adjusted mean eGFR change from baseline to follow-up was 3.3 (95%CI, 1.8-4.8) for empagliflozin *versus* placebo.

Figure 9. Changes in the estimated glomerular filtration rate, based on the TS and measurements up to one day after the last intake of study medication



Source: Packer et al 2020 (109)

Note: Graph shows the adjusted mean changes from baseline in the eGFR as calculated using the Chronic Kidney Disease Epidemiology Collaboration equation. The bars indicate the standard error. The on-treatment data were analysed with a mixed model for repeated measures (MMRM). Age and baseline eGFR were included as linear covariates, while sex, region, baseline LVEF, baseline diabetes status, last projected visit based on dates of randomisation and trial closure, baseline eGFR according to visit, and visit according to treatment interactions were included as fixed effects. TS, treated set.

Thus, when measurements of renal function were compared at the start and after the discontinuation of empagliflozin and placebo, the eGFR declined significantly more in the placebo group than in the empagliflozin group, leading to increased risk of serious renal outcomes, as described in more detail in section B.2.6.2.3. The initial dip in eGFR seen at the start of the treatment with empagliflozin represents a reversible functional change in intrarenal haemodynamics commonly observed with SGLT2is and is not associated with an excess risk of investigator-reported acute kidney injury (111).

B.2.6.2.3 Time to composite renal outcome

Time to the first event in the composite renal endpoint, comprising chronic dialysis, renal transplant or sustained reduction in eGFR (CKD-EPI)cr is shown in Figure 10. The composite renal endpoint occurred in 30 patients (1.6%) in the empagliflozin group and 58 patients (3.1%) in the placebo group, with the sustained reduction in eGFR from baseline of ≥40% being the first recorded renal event in most patients (Table 20). The risk of the composite renal endpoint was halved with empagliflozin compared to placebo (HR, 0.50; 95%CI, 0.32 - 0.77; nominal p = 0.002). Addition of

empagliflozin has therefore demonstrated a large, clinically meaningful benefit in preventing serious renal outcomes in HFrEF patients compared to SoC alone.

Table 20. Cox regression analysis of time to first renal event¶, RS

Time to composite renal outcome*	Placebo (N=1867)	Empagliflozin (N=1863)
Patients with the composite renal endpoint, N (%)	[redacted]	[redacted]
Sustained eGFR reduction ≥40% as the first event	[redacted]	[redacted]
Sustained eGFR <15 mL/min/1.73 m^2 (baseline ≥30) or <10 mL/min/1.73 m^2 (baseline <30) as the first event	[redacted]	[redacted]
Chronic dialysis as the 1st event	[redacted]	[redacted]
Renal transplant as the 1st event	[redacted]	[redacted]
Incidence rate per 100 years at risk	[redacted]	[redacted]
Hazard ratio *vs.* placebo (95% CI), composite renal outcome	0.50 (0.32 - 0.77)	
Nominal p-value	0.0019	

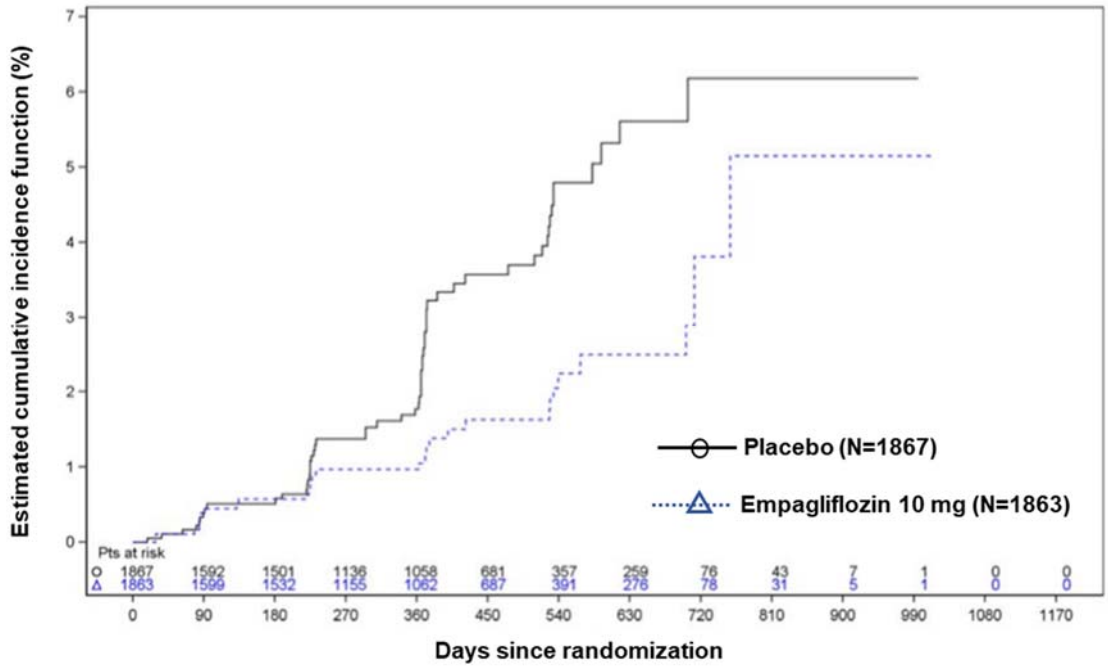
Source: EMPEROR-Reduced CSR, Table 11.1.2.6:1 (135)

Abbreviations: eGFR, estimated glomerular filtration rate; CI, confidence interval; RS, randomised set.

*The composite renal endpoint was comprised of chronic dialysis (with a frequency of twice per week or more for at least 90 days), renal transplant, sustained reduction in eGFR from baseline of ≥40%, sustained eGFR <15 mL/min/1.73m^2 for patients with baseline eGFR ≥30 mL/min/1.73m^2, or sustained eGFR <10 mL/min/1.73m^2 for patients with baseline eGFR <30mL/min/1.73m^2. Sustained was determined by two or more consecutive post-baseline central laboratory measurements separated by at least 30 days (the first to last of the consecutive eGFR values).

¶Cox regression model included covariates age, baseline eGFR (CKD-EPI)cr, region, baseline diabetes status, sex, baseline LVEF, and treatment.

Figure 10. Estimated cumulative incidence function for time to the first event of the composite renal endpoint, RS



Source: EMPEROR-Reduced CSR, Figure 11.1.2.6.1:1 (135)
Note: RS, randomised set.

B.2.6.2.4 Time to first adjudicated hospitalisation for heart failure

Over the duration of the trial, fewer patients experienced the event of first adjudicated HHF in the empagliflozin group ([redacted] %) compared to placebo group ([redacted] The estimated cumulative incidence of first adjudicated HHF, considering all-cause mortality as a competing risk, started to diverge between empagliflozin and placebo groups shortly after randomisation and continued to separate over the course of the trial (Figure 11). The risk of adjudicated HHF was significantly reduced with empagliflozin treatment *versus* placebo (HR, [redacted], [redacted] as determined by the Cox regression model adjusted for age, baseline eGFR (CKD-EPI)cr, region, baseline diabetes status, gender, baseline LVEF, and treatment.

Figure 11. Estimated cumulative incidence function for time to the first adjudicated HHF with all-cause mortality as a competing risk, RS

[redacted]

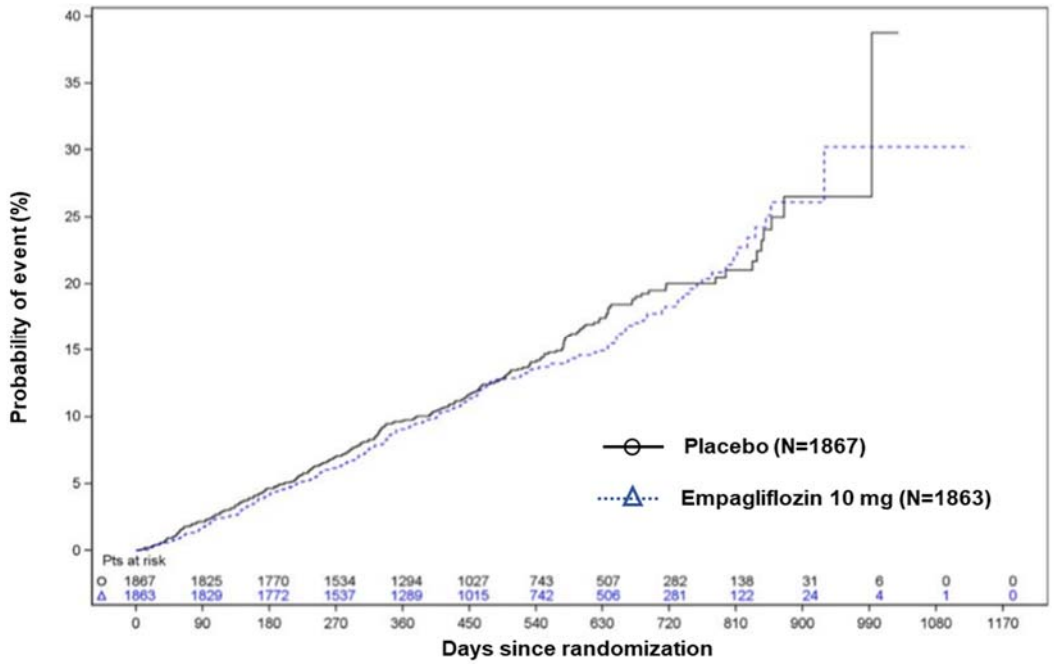
Source: EMPEROR-Reduced CSR, Figure 11.1.2.3.1:1 (135)

B.2.6.2.5 All-cause mortality

The Kaplan-Meier estimate of time to all-cause mortality in the randomised set is shown in

Figure 12. Death from any cause occurred in 249 patients (13.4%) in the empagliflozin group and 266 patients (14.2%) in the placebo group. Cox regression of time to all-cause mortality data for all randomised patients showed that the risk of death from any cause was 8% lower with empagliflozin than with placebo (HR 0.92, 95% CI 0.77-1.10), although the difference did not reach statistical significance (p=0.35).

Figure 12. Time to all-cause mortality, Kaplan-Meier estimate in RS



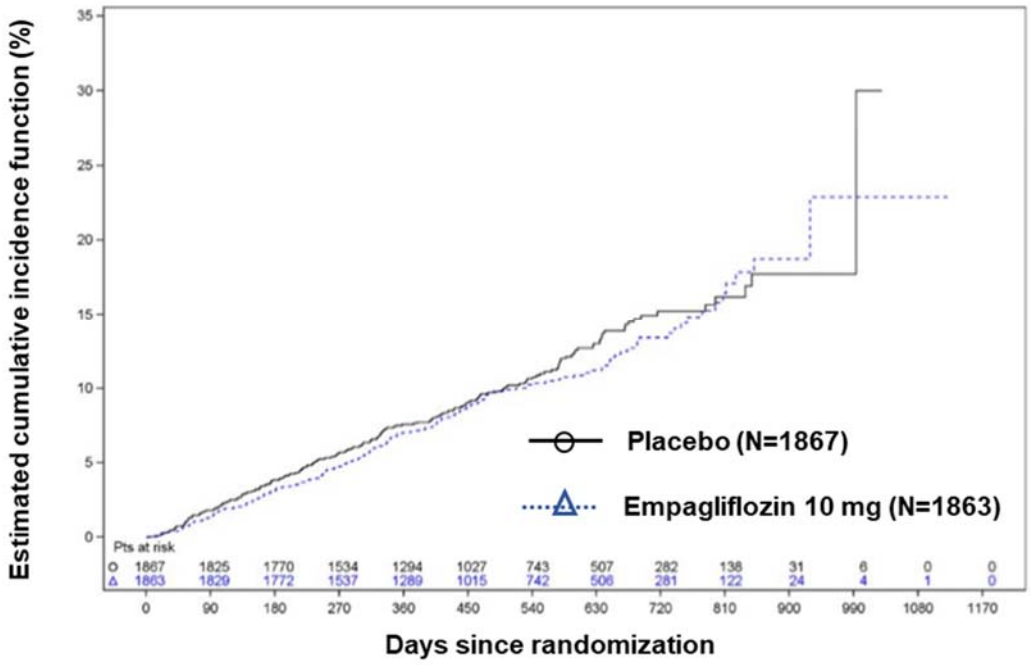
Source: EMPEROR-Reduced CSR, Figure 11.1.2.4.1:1 (135)

Note: RS, randomised set.

B.2.6.2.6 Cardiovascular mortality

Most of the deaths recorded during the study were due to CV causes, such as sudden cardiac death or HF death. Adjudicated CV death occurred in 187 patients (10.0%) in the empagliflozin group and 202 patients (10.8%) in the placebo group. The risk of CV death was 8% lower with empagliflozin relative to placebo (HR, 0.92; 95%CI, 0.75 - 1.12), a difference that did not reach statistical significance (p=0.41). The cumulative incidence of adjudicated CV death in randomised patients, considering non-CV death as a competing risk, is shown in Figure 13.

Figure 13. Estimated cumulative incidence function for time to adjudicated CV death, considering non-CV death as a competing risk- RS

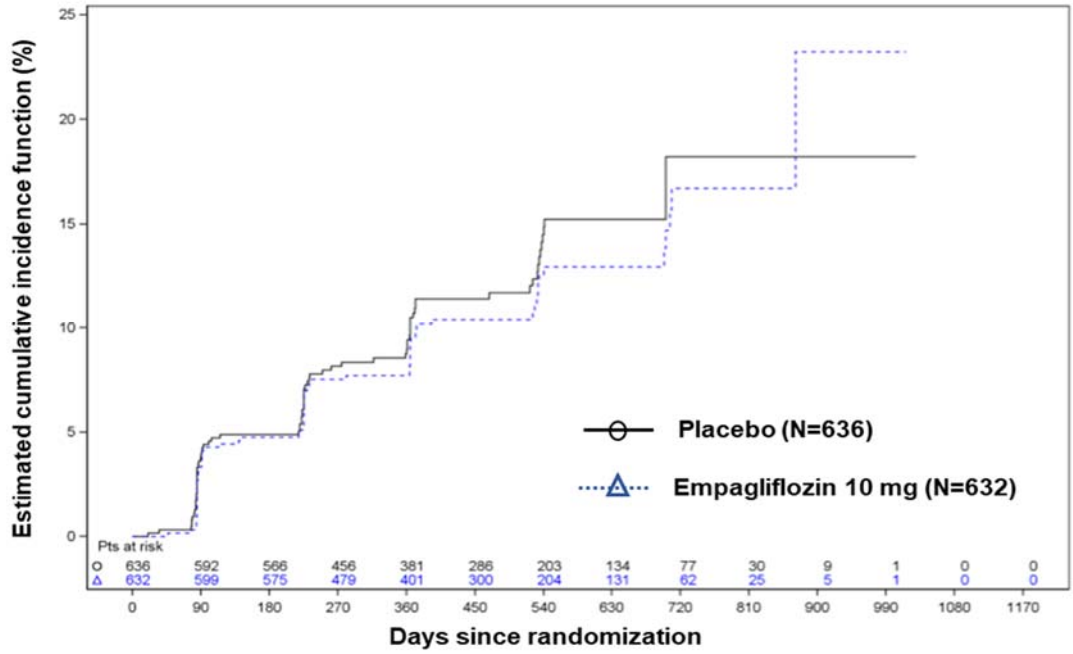


Source: EMPEROR-Reduced CSR, Figure 11.1.2.4.2:1 (135)
Note: RS, randomised set.

B.2.6.2.7 Time to onset of diabetes mellitus (DM) in patients with pre-DM

The onset of DM in patients with pre-DM occurred in 71 of 632 patients in the empagliflozin group (11.2%) and 80 of 636 patients (12.6%) in the placebo group. The observed reduction in risk of onset of DM with empagliflozin compared to placebo (HR, 0.86; 95%CI, 0.62-1.19) was not statistically significant. The estimated cumulative incidence of time to onset of DM in patients with pre-DM, considering all-cause mortality as a competing risk, started to diverge after approximately 8 months, and continued to separate over the remainder of the trial (Figure 14).

Figure 14. Estimated cumulative incidence function for time to onset of DM in patients with baseline pre-DM in the RS



Source: EMPEROR-Reduced CSR, Figure 11.1.2.8.1:1 (135); Note: RS, randomised set.

B.2.6.2.8 First and recurrent all-cause hospitalisation

All-cause hospitalisation occurred in [redacted]) of patients in the empagliflozin group and [redacted]) in the placebo group. The total number of hospitalisation events was lower in the empagliflozin group (1364) than in the placebo group (1570). Analysis of this endpoint using a joint frailty model that accounts for the dependence between recurrent all-cause hospitalisation and all-cause mortality demonstrated that the risk of recurrent all-cause hospitalisation was reduced with empagliflozin treatment compared to placebo ([redacted]). The mean cumulative incidence curves of all-cause hospitalisation in empagliflozin and placebo groups diverged soon after randomisation and maintained

their separation throughout the study (Figure 15). Cox regression showed 18% reduction in risk of first all-cause hospitalisation with empagliflozin compared to placebo (HR, 0.82; 95%CI, 0.74 - 0.90; p<0.0001).

Figure 15. Mean cumulative function for occurrence of all-cause hospitalisation (first and recurrent), RS

[redacted]

Source: EMPEROR-Reduced CSR, Figure 11.1.2.5.1:1 (135)
Note: RS, randomised set

B.2.6.2.9 Further secondary clinical endpoints

Results of further exploratory secondary endpoints from EMPEROR-Reduced trial, including measurement of health status by Kansas City Cardiomyopathy Questionnaire (KCCQ), are presented in Table 21. The change in KCCQ-CSS over time is reported in Figure 20.

Table 21. Summary of further exploratory secondary endpoints from EMPEROR-Reduced study

Endpoint	Placebo (N=1867)	Empagliflozin 10 mg (N=1863)
Time to adjudicated MI (fatal or non-fatal), RS		
Patients with MI, N (%)	[redacted]	[redacted]
Incidence rate per 100 years at risk	[redacted]	[redacted]
HR vs placebo (95% CI)	[redacted]	
Nominal p-value	[redacted]	
Time to adjudicated stroke (fatal or non-fatal), RS		
Patients with stroke, N (%)	[redacted]	[redacted]

Endpoint	Placebo (N=1867)	Empagliflozin 10 mg (N=1863)
Ischaemic	[redacted]	[redacted]
Haemorrhagic	[redacted]	[redacted]
Unclassified	[redacted]	[redacted]
Incidence rate per 100 years at risk	[redacted]	[redacted]
HR vs placebo (95% CI)	[redacted]	
Nominal p-value	[redacted]	
Time to new onset of Afib, as ECG finding or as AE, RS		
Patients without baseline or history of Afib¶	[redacted]	[redacted]
Patients with new onset of Afib, N (%)	[redacted]	[redacted]
Incidence rate per 100 years at risk	[redacted]	[redacted]
HR vs placebo (95% CI)	[redacted]	
Nominal p-value	[redacted]	
Blood pressure changes from baseline to week 52 (mm Hg), RS		
Systolic blood pressure change (mm Hg)	[redacted]	[redacted]
Adjusted mean change from baseline (95% CI)	[redacted]	
p-value	[redacted]	
Diastolic blood pressure change (mm Hg)	[redacted]	[redacted]
Adjusted mean change from baseline (95% CI)	[redacted]	
p-value	[redacted]	
HbA1c (%) change from baseline at week 52, RS patients with diabetes		
Adjusted mean change from baseline	[redacted]	[redacted]

Endpoint	Placebo (N=1867)	Empagliflozin 10 mg (N=1863)
Absolute difference (95%CI)	[redacted]	
QoL measured by KCCQ at 52 weeks§, TS		
Change in clinical summary score at 52 weeks	[redacted]	[redacted]
Adjusted mean change from baseline (95% CI)	[redacted]	
Nominal p-value	[redacted]	
Change in overall summary score at 52 weeks	[redacted]	[redacted]
Adjusted mean change from baseline (95% CI)	[redacted]	
Nominal p-value	[redacted]	
Change in total symptom score at 52 weeks	[redacted]	[redacted]
Adjusted mean change from baseline (95% CI)	[redacted]	
Nominal p-value	[redacted]	

Source: EMPEROR-Reduced CSR, Table 11.1.3.1:1, Table 15.2.4.2.1, Sections 11.1.2.7, 11.1.2.8.2 and 11.1.3.4 (135)

Abbreviations: AE, adverse event; Afib, atrial fibrillation; CI, confidence interval; ECG, electrocardiogram; HbA1c, glycated haemoglobin; HR, hazard ratio; KCCQ, Kansas City Cardiomyopathy Questionnaire; MI, myocardial infarction; QoL, quality of life; RS, randomised set; TS, treated set.

Note: Plus-minus values are means ± SE. Estimates of effect size for time to event endpoints (HR, 95% CI) were derived for the randomised set using Cox regression model which included covariates age, baseline eGFR (CKD-EPI)cr, region, baseline diabetes status, sex, baseline LVEF, and treatment. Continuous endpoints (blood pressure, KCCQ scores) were analysed using mixed model for repeated measures (MMRM).

¶Based on investigator-reported medical history or baseline ECG

§The clinical summary score on the Kansas City Cardiomyopathy Questionnaire ranges from 0 to 100, with higher scores indicating a better quality of life. Analysis of PRO data with a MMRM was based on the treated set and using on-treatment values only.

Frequency of myocardial infarction and stroke were similar between the two treatment groups. New onset of atrial fibrillation was reported less frequently in the empagliflozin group than the placebo group, although the difference did not reach statistical significance since study was not powered to detect differences in exploratory endpoints. There was no marked change in blood pressure in the empagliflozin group, with a placebo-corrected adjusted mean change at week 52 from baseline of -0.7 mmHg (95%CI, -1.8 to 0.4) for systolic and -0.1 mmHg (95% CI, -0.8 to 0.6) for diastolic blood pressure.

The change from baseline in health status was assessed by the Kansas City Cardiomyopathy Questionnaire clinical summary score (KCCQ-CSS) at week 52. The clinical summary score measures HF symptom frequency, symptom burden, and physical limitations. Empagliflozin significantly improved KCCQ-CSS by 1.94, 1.35, and 1.61 points compared to placebo at 3, 8, and 12 months, respectively ($p< 0.05$ for all) (120). A similar improvement was also observed for the KCCQ total symptom score (KCCQ-TSS) and overall summary score (KCCQ-OSS), which includes the quality of life dimension. There were no relevant differences between the treatment groups with regards to health-related quality of life (HRQoL) as assessed by the EQ-5D-5L questionnaire.

B.2.7 Subgroup analyses

The pre-specified subgroup analyses for the efficacy endpoints of EMPEROR-Reduced were:

- Diabetes at baseline (diabetic and non-diabetic patients) (Appendix E)
- Renal function at baseline (eGFR ≥60 mL/min/1.73 m^2, ≤ 60 mL/min/1.73 m^2) (Appendix E)
- Gender
- Race (White, Black/African-American, other including mixed race)
- BMI (<30 kg/m^2 and ≥30 kg/m^2)
- Age (< 65 years and ≥ 65 years) (Appendix E)
- History of HHF in the last 12 months (Yes/No)
- Cause of HF (Ischaemic or non-ischaemic disease)
- NYHA at baseline (II *versus* III/IV)
- Heart failure physiology (reflected in baseline LVEF and level of NT-pro-BNP) (Appendix E)
- Baseline use of mineralocorticoid receptor antagonist
- Baseline use of angiotensin receptor-neprilysin inhibitors (Appendix E)
- Geographic region (Asia, Europe, Latin America, North America, and other)
- Baseline eGFR [≥90, 60 to <90, 45 to <60, 30 to <45, <30 ml/min/1.73 m^2]

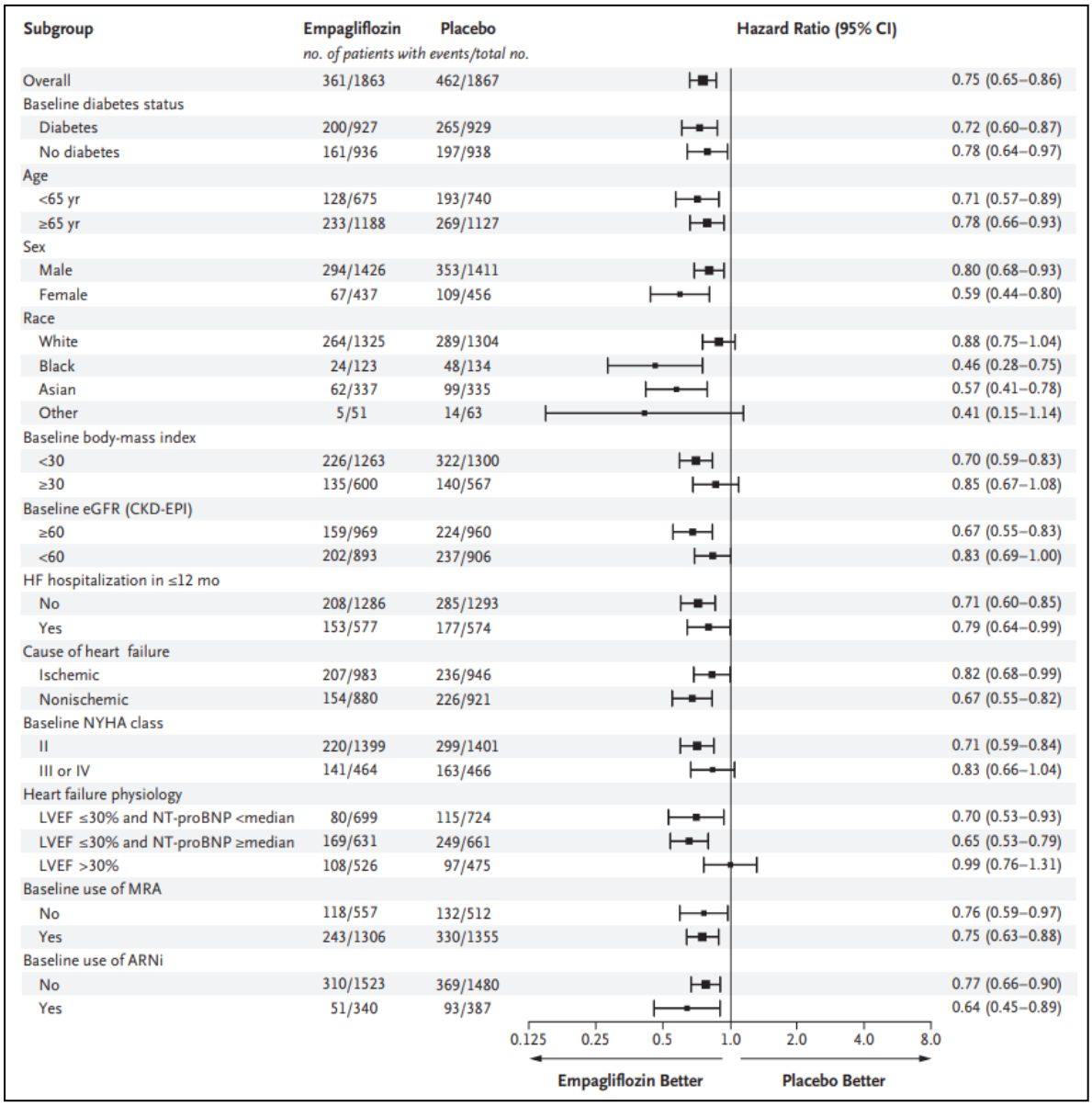
A post-hoc subgroup analysis according to KCCQ-CSS score at baseline (<62.5, 62.5-85.4, ≥85.4) was also conducted. It is worth noting that subgroup analyses were not

adjusted for multiple testing. Hence, the subgroup findings were subject to a greater play of chance and were regarded as hypothesis generating (Appendix E).

The effect of empagliflozin on the combined risk of CV death or HHF was consistent across most pre-specified subgroups, with the point estimate HR less than one in subgroups based on age and gender, BMI, baseline therapies (ARNI/no ARNI; MRA/no MRA), cause of HFrEF, history of HHF, presence or absence of diabetes or impaired renal function (Figure 16). Of note, the consistent effect of empagliflozin in patients with an eGFR lower than 60 ml/min per 1·73m² provides evidence of an important reduction of CV death or HHF in this high-risk subgroup, including patients with eGFR as low as 20 ml/min/1.73 m^2 (111).

Numerical differences were observed in the subgroup analyses by baseline NYHA class and physiology of HF. Although the direction of the treatment effect remained consistent, magnitude of the benefit was smaller in the subgroup with NYHA class III-IV (more severe HF) versus NYHA class II (less severe HF) at baseline. Subgroup analyses by other measures of HF severity however did not support the same directionality of effect since larger effect was seen in a higher severity subgroup (LVEF ≤ 30% and NT-pro-BNP < median) *versus* lower severity subgroup (LVEF>30%), while no variation was seen across KCCQ-CSS tertiles (Appendix E) or NT-pro-BNP tertiles (135). Furthermore, the point estimate HR remained less than one for each subgroup.

Figure 16. Primary outcome of EMPEROR-Reduced in pre-specified subgroups



Source: Packer et al 2020 (109)

Abbreviations: ARNi, angiotensin receptor-neprilysin inhibitor; CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration; HF, heart failure; eGFR, estimated glomerular filtration rate; LVEF, left ventricular ejection fraction; MRA, mineralocorticoid receptor antagonist; NT-pro-BNP, N-terminal prohormone of brain natriuretic peptide; NYHA, New York Heart Association.

Note: The size of the squares for the hazard ratios is proportional to the size of the subgroup. Interaction p values are nominal; the subgroup analyses were not adjusted for multiple testing. The body-mass index is the weight in kilograms divided by the square of the height in meters. Race was reported by the patients.

The use of data from Europe subgroup to assess generalisability is not appropriate and could contribute to existing ethnic inequalities in health (142), contrary to the NICE's Social Value Judgments and the Equality Act 2010 (race is one of the protected characteristics) (14). These data will not be reported separately in this submission. The Europe subgroup of EMPEROR-Reduced was [redacted] white and therefore not representative of the multi-ethnic UK population, which consists of 86% white, 3.3% black, 7.5% Asian and 3.2% other (143). This difference is even wider in

the metropolitan areas of the UK (44.9% white in London) (143). The ITT population of EMPEROR-Reduced, which was 71% white, 6.6% black, 18.1% Asian and 4.2% other (109) is more generalisable to the ethnically diverse UK population and is, therefore, the population considered in the economic analysis. This is consistent with the committee's perspective in the dapagliflozin appraisal (TA679) (4). Although the ERG preferred the European subgroup of DAPA-HF trial for the base-case cost-effectiveness analysis (CEA), the committee recognised that the European subgroup was predominantly white, comprised less than half of the overall trial population and may have an absolute risk of complications different to that of patients from the rest of the world (4). The committee therefore concluded that data from the overall DAPA-HF population were acceptable for decision making.

The use of ITT population for the CEA of empagliflozin in HFrEF is also the most statistically robust approach since EMPEROR-Reduced was not powered to evaluate the treatment effect in subgroups. With many subgroup analyses carried out without adjusting the overall significance level of the trial, it is unclear if the results represent spurious findings. Given that the results for the subgroups were generally consistent with the confirmatory analyses, only the ITT population was considered in the economic analysis. Results of the clinically relevant pre-specified subgroups can be found in Appendix E.

B.2.8 Indirect and mixed treatment comparisons

- Evidence for empagliflozin vs standard treatment with ACEi, ARBs, BB, MRAs is the most relevant for the committee to consider. A comparative analysis of empagliflozin vs dapagliflozin

is presented here on the request of the ERG and NICE technical team at the Decision Problem meeting; although there is limited evidence that dapagliflozin represents usual care. Its market share is [redacted]) as of MQT May 2021 (B.1.1) (5). In the HF only population, it is prescribed[redacted] times less often than sacubitril/valsartan (Table 2) (5). As future use of treatment are speculative, we can only reflect the care pathway used today in this submission. Thus, it would be inappropriate to place emphasis on this comparison for decision making.

- A Bucher indirect treatment comparison reported comparable efficacy of empagliflozin versus dapagliflozin across key outcomes, using placebo as the common comparator arm.
- A Bucher comparison represents the best available evidence. The NICE Guide to Methods 2013 recommends a Bucher ITC unless there is evidence that population adjustment (i.e. a match adjusted indirect comparison (MAIC)) would result in the removal of bias (144). Due to a lack of overlap in the inclusion criteria for NT-proBNP and LVEF, there is limited evidence that a MAIC would reduce bias. This conclusion was also reached in the dapagliflozin technology appraisal for an indirect comparison of dapagliflozin vs sacubitril/valsartan (6). A Bucher ITC was preferred by the committee over a MAIC despite the PARADIGM-HF trial being more dissimilar to DAPA-HF than EMPEROR-Reduced.
- There are several implications for these analyses:
  - The conclusion from the Bucher ITC that empagliflozin and dapagliflozin offer comparable efficacy across key clinical outcomes for patients with HFrEF is consistent with feedback from UK clinical experts.
  - UK clinicians advised that they want to tailor treatment to individual patients; and would value having more than one SGLT2i available (103, 145, 146) This would enable continuity of care, use in patients with an eGFR 20 to 30 mL/min/1.73m^2, and management of specific AEs.
  - As the key efficacy outcomes are comparable, the cost effectiveness of SGLT2i vs SoC is the most relevant economic evidence to consider (B.3.7) and supports a scenario where multiple SGLT2is are recommended for use. This utilizes a pooled meta-analysis by Zannad et al 2020 (119). The pooled meta-analysis has a larger sample size, and thus the ICER for SGLT2i vs SoC provides a more robust estimate of cost estimate than empagliflozin vs SoC or dapagliflozin vs SoC alone. The cost effectiveness of SGLT2i vs SoC is presented in B.3.7 and is consistent with Section 5.1.14 of the NICE Guide to Methods 2013 (8). An economic comparison of empagliflozin vs dapagliflozin is not presented as it will not support prescriber choice nor individualisation of care.

B.2.8.1 Objective of the indirect comparison

The evidence for empagliflozin vs standard treatment with ACEi, ARBs, BB, MRAs, sacubitril/valsartan is the most relevant for the committee to consider as these products represent the mainstay of treatment for HFrEF in the UK. As dapagliflozin is

included in the Final Scope, analyses of the relative efficacy of empagliflozin and dapagliflozin for HFrEF is presented here, however these are supplementary. In the absence of a head-to-head trial, a Bucher ITC was performed to estimate the relative efficacy of empagliflozin *versus* dapagliflozin in adult patients with symptomatic but stable HFrEF using evidence from the EMPEROR-Reduced and DAPA-HF trials, respectively (Table 22). These ITCs are supported by the results of a pooled meta-analysis of SGLT2i vs SoC.

Rationale for analytical approach

A Bucher ITC represents the best available evidence. Technical Support Document 18 (TSD18) [(147)Pg 61 to 63] recommends a Bucher ITC unless there is evidence that differences in the patient populations across trials might introduce bias and effect the estimate of relative efficacy. When there are differences, alternatives such as an MAIC should be considered if they can adequately remove bias. For this comparison, a MAIC is not appropriate because:

- UK-based clinical experts have fed back that in clinical practice both dapagliflozin and empagliflozin are considered comparable for the treatment of HFrEF. It is biologically plausible that they are comparable given that they belong to the same drug class.
- As a MAIC relies on their being sufficient overlap, it's unlikely that population adjustment would remove any bias arising from differences in the inclusion criteria between EMPEROR-Reduced and DAPA-HF for LVEF and NT-pro-BNP.

Table 22.Summary of the trials used to carry out the indirect treatment comparison

Trial reference	EMPEROR-Reduced	DAPA-HF

Intervention (N)	Empagliflozin (10 mg qd) + SoC	Dapagliflozin (10 mg or 5 mg qd) + SoC
Comparator (N)	Placebo + SoC	Placebo + SoC
Study start completion (years)	2017–2020	2017–2019
Phase	III	III
Method of blinding	Double-blind	Double-blind
Randomisation	1:1, stratified by geographical region, history of diabetes and eGFR	1:1, stratified by type II diabetes (with and without)
Study centres	Multicentre (Europe, North America, Latin America, Asia, Other)	Multicentre (Europe, North America, Latin America, Asia Pacific)
Primary composite	The composite primary endpoint for this trial was the time to first event of adjudicated CV death or adjudicated HHF	Time to the first occurrence of any of either CV death, hospitalisation for HF or an urgent HF visit
Secondary outcomes	Key secondary outcomes: • Occurrence of adjudicated HHF (first and recurrent) • eGFR $(CKD\text{-}EPI)_{cr}$ slope of change from baseline Other secondary outcomes: • Time to the first event in the composite renal endpoint: chronic dialysis, renal transplant, or sustained reduction in eGFR $(CKD\text{-}EPI)_{cr}$ • Time to first adjudicated HHF • Time to adjudicated CV death • Time to all-cause mortality • Time to onset of T2DM in patients with pre-T2DM • Change from baseline in KCCQ clinical summary at week 52 • Occurrence of all-cause hospitalisation (first and recurrent)	• Time to the first occurrence of CV death or hospitalisation for HF • Total number of (first and recurrent) HF hospitalisations and CV death • Change from baseline measured at 8 months in KCCQ overall summary score • Renal composite: ≥50% sustained decline in eGFR, reaching end-stage renal disease or renal death • Time to death from any cause
Median follow-up duration	16 months	18.2 months

Abbreviations: N, number of participants; qd, once a day; SoC, standard of care.

B.2.8.2 Evidence base and comparators

A clinical SLR was conducted to identify all relevant randomised controlled trial evidence related to the treatment of HFrEF (Appendix D). The SLR identified a total of 45 studies from 356 publications, including three studies that reported outcomes with empagliflozin (109, 119, 120) and four studies describing efficacy outcomes with dapagliflozin (148-151). The final evidence base for the ITC considered the primary studies for empagliflozin versus placebo as an add-on to SoC (EMPEROR-Reduced) (109) and dapagliflozin versus placebo as an add-on to SoC (DAPA-HF) (150) which were the pivotal phase III trials informing evidence requirements for the regulatory approval of these interventions (Table 22).

B.2.8.3 Results of the Bucher ITCs and pooled meta-analysis

The Bucher ITCs reported comparable efficacy of empagliflozin *versus* dapagliflozin, using placebo as the common comparator arm. The strongest evidence on the similarity of empagliflozin versus dapagliflozin comes from an ITC of the primary composite endpoint (adjudicated CV death or HHF). When the DAPA-HF primary composite was compared to the EMPEROR-Reduced primary composite, the HR was not significant ([redacted]). The primary composite endpoints across EMPEROR-Reduced and DAPA-HF were defined slightly differently (Table 22). Even when the outcome was defined as per the EMPEROR-Reduced primary composite (time to first event of adjudicated CV death or adjudicated HHF); comparable efficacy was observed ([redacted]) (Table 23).

A trend for comparable efficacy was observed for key secondary outcomes, including time to first HHF, total HHF and KCCQ; with HRs of approximately 1 and overlapping confidence intervals (Table 23). The Bucher ITC of worsening renal function, as defined in DAPA-HF, yielded the HR (and 95% CI) of 0.73 (0.34, 1.56), suggesting that empagliflozin might be slightly more effective in reducing the hazard of worsening renal function relative to dapagliflozin, although the confidence interval was wide and contained the no difference value of 1.0.

The Bucher ITCs also showed comparable efficacy for both CV death and all-cause mortality (HR [empagliflozin vs dapagliflozin] [redacted]), respectively) (Table 23). A numerically lower death rate was observed for dapagliflozin than empagliflozin; however, this should be interpreted in context. In EMPEROR-Reduced, both time to

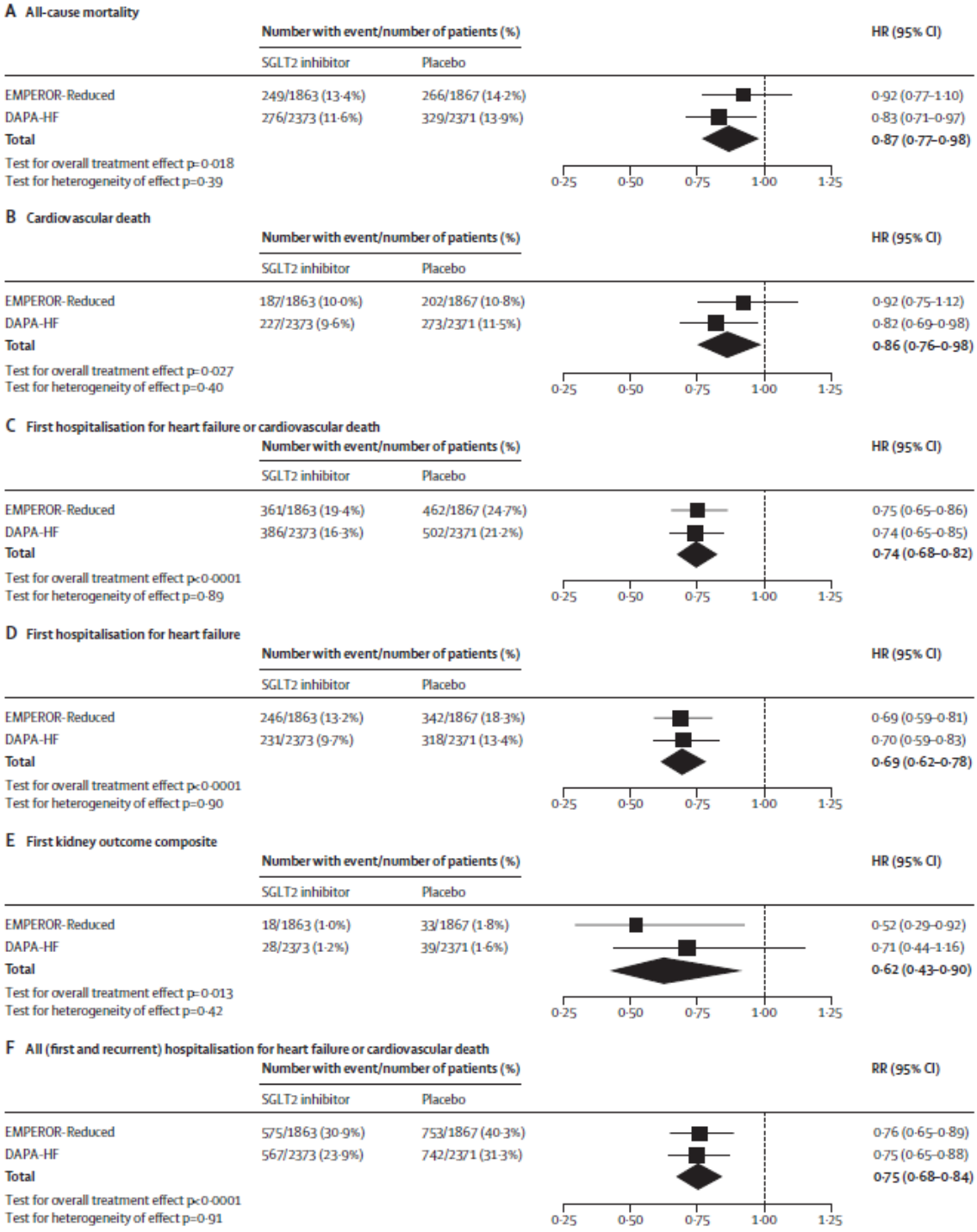
CV-death and all-cause mortality were included as a standalone secondary endpoint. In DAPA-HF, neither of these were listed individually as a secondary or exploratory endpoint. In both trials, the sample was not large enough to show a statistically significant difference with sufficient power between intervention and placebo for these outcomes individually. The sample size was based on a patient experiencing one of the composite events for the primary composite, which included terminal and non-terminal events. In EMPEROR-Reduced, 841 primary composite events and 2850 patients were needed to achieve a statistical power of 90% (assumptions: HR 0.8, 15% annual CV death event rate for placebo, 2-sided at α=0.05, 18 months accrual and 20 months follow-up)(135). In the DAPA-HF trial, 844 primary composite events and 4500 patients were needed to provide a statistical power of 90% (assumptions: HR 0.8, 11% annual CV death event rate for placebo, 1-sided at α=0.01, 18 months accrual and 24 months follow-up)(150).

A larger sample size, accrual and follow-up time increases the certainty that the point estimate for an outcome is the true effect size in a population. Conversely, with a smaller sample size, there is greater variability in the point estimate, and this is observed in wider confidence intervals. Due to a smaller sample size, shorter accrual and follow-up time in EMPEROR-Reduced than in DAPA-HF, there was greater uncertainty in where the true effect size for CV death and all-cause mortality lay in the population. This is evident in there being wider confidence intervals for the HR for CV death being in EMPEROR-Reduced than DAPA-HF (EMPEROR-Reduced, 0.92 (0.75-1.12) versus DAPA-HF, 0.82 (0.69-0.98)) (109, 150). In a fixed effects pooled meta-analysis of EMPEROR-Reduced and DAPA-HF with a combined sample size of 8,474, the variability in the CI reduced to 22% from 37% and 29% in EMPEROR-Reduced and DAPA-HF, respectively (119). For the pooled analysis, the HR for CV death for SGLT2i vs placebo for the ITT population was 0.86 (0.76-0.98); for all-cause mortality it was 0.87 (0.77-0.98). This pooled analysis increased the certainty in the estimate of the true effect size in the population (Table 23) (119).

Table 23. Summary of Bucher ITC results for empagliflozin plus SoC versus dapagliflozin plus SoC (EMPEROR-Reduced vs DAPA-HF, ITT population)

Endpoint: relative effect measure	EMPEROR-REDUCED: empagliflozin versus placebo[a]	DAPA-HF: dapagliflozin versus placebo[a]	Bucher ITC: empagliflozin versus dapagliflozin[a]
Time to first event of adjudicated CV death or adjudicated HHF: HR (95% CI)	0.75 (0.65, 0.86)	0.75 (0.65, 0.85)	[redacted]
Time to first event of adjudicated CV death or adjudicated HHF (EMPEROR-Reduced) versus Time to first worsening of heart failure (hospitalisation or an urgent visit resulting in intravenous therapy for heart failure) or CV death (DAPA-HF) : HR (95% CI)	0.75 (0.65, 0.86)	0.74 (0.65, 0.85)	[redacted]
Time to first adjudicated HHF: HR (95% CI)	0.69 (0.59, 0.81)	0.70 (0.59, 0.83)	[redacted]
Time to adjudicated CV death: HR (95% CI)	0.92 (0.75, 1.12)	0.82 (0.69, 0.98)	[redacted]
Time to all-cause mortality: HR (95% CI)	0.92 (0.77, 1.1)	0.83 (0.71, 0.97)	[redacted]
Occurrence of adjudicated HHF (first and recurrent) – analysed using a joint frailty model: HR (95% CI)	0.70 (0.58, 0.85)	0.71 (0.61, 0.82)	[redacted]
Occurrence of adjudicated HHF (first and recurrent) – analysed using a Lin-Wei-Yang-Ying model: RR (95% CI)	0.76 (0.65, 0.89)	0.75 (0.65, 0.88)	[redacted]
Worsening renal function (as defined in DAPA-HF): HR (95% CI)	0.52 (0.29, 0.92)	0.71 (0.44, 1.16)	[redacted]
Change in KCCQ total symptom score at 8 months/7.4 months: MD (SE/95% CI)	1.6 (0.7)	2.8 (0.5)	[redacted]

Figure 17. Meta-analysis of EMPEROR-Redcuced and DAPA-HF trials (ITT population)



B.2.8.4 Study heterogeneity

As stated in Section B.2.8.1, unless there is evidence of treatment effect modifiers, a Bucher ITC is recommended by NICE. The Bucher ITC and pooled fixed effects meta-analysis rely on the assumption of homogeneity. That is, there are no differences in the distribution of patient populations between EMPEROR-Reduced and DAPA-HF that impact the estimate of relative efficacy; i.e. no treatment effect modifiers. This assumption allows the combination of their relative effects. If this assumption does not hold, a population adjusted ITC (such as a MAIC) should be considered. Consistent with the NICE Guide to Methods 2013 (Section 5.2.7 to 5.2.11) (8), an assessment of heterogeneity and investigation into the impact of treatment effect modifiers has been undertaken.

B.2.8.4.1 Trial design

Both studies were phase III, multicentre, double-blind RCT, conducted over the same period in similar geographical locations suggestive of consistent clinical practices across both trials. Compared with those in DAPA-HF, patients enrolled in the EMPEROR-Reduced were more likely to have been treated with ARNI at baseline (20% vs. 11%) and to have received implantable defibrillator (32% vs. 26%) or resynchronisation therapy (12% vs. 7%) (Table 24). Using the placebo arms of the EMPEROR-Reduced and DAPA-HF as a common comparator assumes that all treatments comprising SoC have equivalent efficacy and the differences in the proportion of patients receiving each type of treatment have no impact on the relative efficacy; i.e. background treatment is not an effect modifier. In a post-hoc analysis of the DAPA-HF trial, Docherty et al. 2020 (152) found that the relative efficacy of dapagliflozin versus placebo was consistent across the following yes/no subgroups: diuretic, digoxin, MRA, sacubitril + valsartan, ivabradine, implanted cardioverter-defibrillating device, and cardiac resynchronisation therapy. Similarly, there was no significant difference in relative efficacy of empagliflozin versus placebo in patients receiving or not receiving ARNI (112). The use of the placebo arm of EMPEROR-Reduced and DAPA-HF as a common comparator in the Bucher ITC and for the fixed effects pooled meta-analysis is justified.

Table 24. Standard of care received at baseline in EMPEROR-Reduced and DAPA-HF trials

SoC at baseline	EMPEROR-Reduced		DAPA-HF	
	Empagliflozin	Placebo	Dapagliflozin	Placebo
Diuretic	94.2%	95.9%	93.4%	93.5%
ACE inhibitor or ARB	70.5%	68.9%	84.2%	82.4%
Sacubitril valsartan	18.3%	20.7%	10.5%	10.9%
Beta-blocker	94.7%	94.7%	96.0%	96.2%
MRA	70.1%	72.6%	71.5%	70.6%
Digitalis	15.2%	16.7%	18.8%	18.6%
Implantable cardioverter-defibrillator	31.0%	31.8%	26.2%	26.1%
Cardiac resynchronisation therapy	11.8%	11.9%	8.0%	6.9%

Abbreviations: ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blocker; MRA, mineralocorticoid receptor antagonist.

B.2.8.4.2 Inclusion and exclusion criteria

EMPEROR-Reduced and DAPA-HF included patients with chronic HFrEF and LVEF≤ 40% (although the timing of the measurements differed between these two studies) (109, 150). In EMPEROR-Reduced, the NT-pro-BNP inclusion criteria differed depending on the reduction in EF, prior HHF and whether the patient had atrial fibrillation (varying from ≥ 600 pg/ml to ≥ 5,000 pg/ml). There were no specific cut-offs in DAPA-HF (Table 25). Based on these differences the target population of EMPEROR-Reduced included patients with mild, moderate and severe disease while DAPA-HF included patients with mild to moderate disease.

Table 25. Summary of key inclusion/exclusion criteria

	EMPEROR-Reduced	DAPA-HF
NYHA class	II-IV	II-IV (present for at least 2 months)
Reduced EF	≤ 40% (measured within 6 months prior to Visit 1 or after study consent)	≤ 40% (within last 12 months prior to enrolment)
NT-pro-BNP	EF ≥ 36% to ≤ 40%: WO AF: ≥ 2,500 pg/mL AF: ≥ 5,000 pg/mL EF ≥ 31% to ≤ 35%: WO AF: ≥ 1,000 pg/mL AF: ≥ 2,000 pg/mL EF ≤ 30%: WO AF: ≥ 600 pg/mL AF: ≥ 1,200 pg/mL For EF ≤ 40% and HHF within previous 12 months: WO AF: ≥ 600 pg/mL AF: ≥ 1,200 pg/mL	WO AF: ≥ 600 pg/mL ≥ 400 pg/ml if HHF within previous 12 months AF: ≥ 900 pg/mL
Prior HHF	NA	NA

Abbreviations: AF, atrial fibrillation; EF, ejection fraction; HF, heart failure; HHF, hospitalisation for heart failure; NT-pro-BNP, N-terminal pro hormone B-type natriuretic peptide; NA, not applicable; NYHA, New York Heart Association; WO, without.

B.2.8.4.3 Baseline characteristics

The baseline characteristics between EMPEROR-Reduced and DAPA-HF were broadly comparable across several variables, including age, sex, region, prior Ischaemic HF and atrial fibrillation. However, there were some differences. The EMPEROR-Reduced trial was enriched with sicker patients compared to DAPA-HF, based on LVEF and NT-pro-BNP. Baseline LVEF was lower for (27% vs 31%) and the median NT-pro-BNP levels higher (~1900pg/mL vs 1430 pg/mL) in EMPEROR-Reduced compared to DAPA-HF. However, fewer patients in EMPEROR-Reduced vs DAPA-HF were hospitalised for HHF in the preceding 12 months (~31% vs 47%), and numerically more patients in EMPEROR-Reduced were in NYHA class II. Baseline eGFR was lower in EMPEROR-Reduced (~61 ml/min/1.73m2) compared to DAPA-HF (~66 ml/min/1.73m^2) (Table 26).

Table 26. Comparison of baseline characteristics of subjects enrolled in EMPEROR-Reduced and DAPA-HF trials

	EMPEROR-Reduced		DAPA-HF	
Treatment (N)	Empagliflozin (N = 1,863)	Placebo (N = 1,867)	Dapagliflozin (N = 2,373)	Placebo (N = 2,371)
Age, mean (SD)	67.2 (10.8)	66.5 (11.2)	66.2 (11.0)	66.5 (10.8)
Female sex, n (%)	437 (23.5)	456 (24.4)	564 (23.8)	545 (23.0)
North America, n (%)	212 (11.4)	213 (11.4)	335 (14.1)	342 (14.4)
South/Latin America, n (%)	641 (34.4)	645 (34.5)	401 (16.9)	416 (17.5)
Europe, n (%)	676 (36.3)	677 (36.3)	1,094 (46.1)	1,060 (44.7)
Asia Pacific, n (%)	248 (13.3)	245 (13.1)	543 (22.9)	553 (23.3)
NYHA I, n (%)	0	0	0	0
NYHA II, n (%)	1399 (75.1)	1401 (75.0)	1,606 (67.7)	1,597 (67.4)
NYHA III, n (%)	455 (24.4)	455 (24.4)	747 (31.5)	751 (31.7)
NYHA IV, n (%)	9 (0.5)	11 (0.6)	20 (0.8)	23 (1.0)
LVEF – %, mean (SD)	27.7 (6.0)	27.2 (6.1)	31.2 (6.7)	30.9 (6.9)
NT-pro-BNP – pg/ml, median (IQR)	1,887 (1077, 3429)	1,926 (1153, 3525)	1,428 (857, 2,655)	1,446 (857, 2,641)
Ischaemic HF, n (%)	983 (52.8)	946 (50.7)	1316 (55.5)	1358 (57.3)
HHF, n (%)	577¶ (31.0)	574¶ (30.7)	1,124 (47.4)	1,127 (47.5)
Atrial fibrillation, n (%)	664 (35.6)	705 (37.8)	916 (38.6)	902 (38.0)
Diabetes mellitus, n (%)	927 (49.8)	929 (49.8)	993 (41.8)	990 (41.8)

	EMPEROR-Reduced		DAPA-HF	
eGFR – ml/min/1.73m^2 Mean (SD)	61.8 (21.7)	62.2 (21.5)	66.0 (19.6)	65.5 (19.3)

Abbreviations: AF, atrial fibrillation; eGFR, estimated glomerular filtration rate; HF, heart failure; HHF, hospitalisation for heart failure; LVEF, left ventricular ejection fraction; NT-pro-BNP, N-terminal pro hormone B-type natriuretic peptide; NYHA, New York Heart Association. Rows highlighted in pink describe baseline characteristic with significant variation between the two trials.
¶In EMPEROR-Reduced, the number of HHF refers to the previous 12 months, while there was no time limit on prior HHF in DAPA-HF.

B.2.8.4.4 Outcomes

The outcomes included in EMPEROR-Reduced and DAPA-HF are summarised in Table 27. There were differences in how the primary composite endpoint was defined across trials (described in B.2.8.3) and key secondary endpoints.

The definition of the composite renal outcome in DAPA-HF (time to the first occurrence of any of the components of the composite: ≥ 50% sustained decline in eGFR or reaching end-stage renal disease (ESRD) or renal death') was different to that in EMPEROR-Reduced. Sufficient data was available to re-estimate the HR in EMPEROR-Reduced based on the definition of this endpoint in DAPA-HF and therefore a comparison for this endpoint was possible. The KCCQ total symptom score was reported at different time points in the two trials. The analysis of patient-level data from EMPEROR-Reduced allowed the comparison of change from baseline in KCCQ total symptom score to be made at a similar time point in both trials, week 32 in EMPEROR-Reduced and at 8 months (≈ 35 weeks) in DAPA-HF. The slope of change in eGFR (CKD-EPI)cr from baseline, time to onset of DM, and occurrence of all-cause hospitalisation (first and recurrent) were not reported in DAPA-HF.

Table 27. Availability of EMPEROR-Reduced primary and secondary endpoints in DAPA-HF trial, ITT population

Primary/secondary endpoints in EMPEROR-Reduced	DAPA-HF
Time to first event of adjudicated CV death or adjudicated HHF in patients with HFrEF	✓
Occurrence of adjudicated HHF (first and recurrent)	✓
eGFR $(CKD\text{-}EPI)_{cr}$ slope of change from baseline	×

Primary/secondary endpoints in EMPEROR-Reduced	DAPA-HF
Time to first occurrence of chronic dialysis or renal transplant or sustained reduction of ≥ 40% eGFR $(CKD\text{-}EPI)_{cr}$	×
Time to first adjudicated HHF	✓
Time to adjudicated CV death	✓
Time to all-cause mortality	✓
Time to onset of DM	×
Change from baseline in KCCQ-CSS at week 52	×[#]
Occurrence of all-cause hospitalisation (first and recurrent)	×

Abbreviations: ARNI, angiotensin receptor neprilysin inhibitor; CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration equation; cr, creatinine; CV, cardiovascular; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; HHF, hospitalisation for heart failure; HFrEF, heart failure with reduced ejection fraction; ITT, intention to treat; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire Clinical Summary Score.
[¶]Similar endpoint, time to the first occurrence of any of the components of the composite ≥50% sustained decline in eGFR or reaching ESRD or renal death, was reported.
[#]Similar endpoint, change in total symptom score at 8 months, was reported.

B.2.8.4.5 Treatment effect modifiers

Bucher ITCs

Bucher ITCs represents the best available evidence to address the comparison requested by NICE. There is limited evidence that alternative approaches, such as MAICs, would adequately address any heterogeneity that might impact the estimate of relative treatment effect.

Treatment effect modifiers were identified through exploration of differences in treatment effect within pre-specified subgroup analysis, clinical validation and patient-level data analysis for variables not already assessed in the EMPEROR-Reduced clinical trial report.

Potential treatment effect modifiers for time to event outcomes are listed in Table 28. These were race, NYHA, LVEF, NT-pro-BNP, and eGFR. Justification for their inclusion is as follows:

- **Race:** There were numerical differences in HRs for white vs non-white patient in a subgroup analysis of the primary composite endpoint (CV death or HHF) in EMPEROR-Reduced. White patients displayed a more modest effect of

empagliflozin vs SoC than other races; but all HR were below 1 (Figure 16) (135).

- **NYHA:** A publication of subgroup analysis for the primary composite endpoint for DAPA-HF noted that the effect of dapagliflozin was generally consistent across pre-specified subgroups, although patients in NYHA III or IV appeared to have less benefit (relative to the control arm) than those in class II. All HRs were still below 1, but NYHA III/IV group had wider confidence intervals with the upper limit crossing 1 (Figure 18) (150). NYHA was therefore considered for inclusion in the matching variables for the time to event outcomes.
- **LVEF, NT-pro-BNP and eGFR**: These variables were identified from supplementary analyses of EMPEROR-Reduced patient-level data. Selection was based upon, whether the effect of the interaction between treatments and covariate had a p-value of less than 0.1 and whether the interaction term was selected from a backwards selection process. The backwards selection process started with a model that included interaction terms for all potential treatment effect modifiers.

These were a targeted list of treatment effect modifier. A broader list of treatment effect modifiers was considered, similar to a recent published MAIC analysis comparing dapagliflozin and sacubitril/valsartan, however this list was narrowed down following clinical input. A UK cardiologist confirmed that the identification of these five variables as treatment effect modifiers made clinical sense and that the associated matching categories were clinically meaningful.

The objective of a MAIC is to reduce bias by making adjustments to the study populations at baseline. However, due to the lack of overlap between LVEF and NT-pro-BNP, this is unlikely to be achieved for the comparison of empagliflozin vs dapagliflozin. Baseline LVEF and NT-pro-BNP required the largest adjustment in EMPEROR-Reduced to match the DAPA-HF population (Table 29). To match the median values of baseline characteristics in DAPA-HF, patients in EMPEROR-Reduced with a LVEF [redacted] and those with a median NT-pro-BNP level of [redacted]. Further, the effective sample size for EMPEROR-Reduced would be reduced by [redacted] for both LVEF and NT-pro-BNP respectively. A small effective sample size indicates that the weights applied for a MAIC are highly variable due to a lack of population overlap, and

that the estimate might be unstable. The weighting for a MAIC was applied to LVEF and NT-pro-BNP individually. However, in EMPEROR-Reduced, with an LVEF >32% it is only possible to have a NT-pro-BNP >1000 pg/mL at baseline whereas a patient in DAPA-HF could have a NT-pro-BNP >600 pg/mL. Since a lower NT-pro-BNP level indicates better health; it's likely that sicker patients in EMPEROR-Reduced are being matched to healthier patients in DAPA-HF.

A Bucher ITC was preferred over a MAIC for the technology appraisal for dapagliflozin (TA679)(6) for a comparison comparing dapagliflozin plus SoC versus sacubitril/valsartan plus SoC, even though the PARADIGM-HF trial was more dissimilar to DAPA-HF than EMPEROR-Reduced. Unlike EMPEROR-Reduced, PARADIGM-HF had a five to ten week run in period where patients received enalapril followed by sacubitril valsartan prior to randomisation. The ERG noted that the use of a MAIC was not fully justified. There was no evidence that a population adjustment would result in the removal of bias.

Pooled fixed effect meta-analyses

The pooled meta-analysis reported by Zannad et al 2020(119) was a fixed effects model. Like a Bucher ITC, a pooled fixed model assumes that there is limited heterogeneity in the population and the confidence intervals are not adjusted. Conversely, a random effects model adjusts for heterogeneity through repeated sampling. A random effects model should be interpreted with caution. With only two studies, the degrees of freedom to robustly estimate inter-study variability using a Cochrane Q test is very limited. Thus, these tests should be interpreted with caution.

B.2.8.5 Implications of the indirect comparisons

There are several implications for these analyses:

- The conclusion from the Bucher ITC that empagliflozin and dapagliflozin offer comparable efficacy across key outcomes for patients with HFrEF is consistent with feedback from UK clinical experts.

- UK clinicians advised that they want to tailor treatment to individual patients; and would value having more than one SGLT2i available. (145) (103, 146) This would

enable continuity of care, use in patients with an eGFR 20 to 30 mL/min/1.73m^2, and management of specific AEs.

- As the key efficacy outcomes are comparable, the cost-effectiveness of SGLT2i vs SoC is the most relevant economic evidence to consider (0) and supports a scenario where multiple SGLT2is are recommended for use. This utilises the pooled meta-analysis by Zannad et al 2020(119). A pooled meta-analysis has a larger sample size, thus the ICER for SGLT2i vs SoC provides a more robust estimate of cost estimate than empagliflozin vs SoC or dapagliflozin vs SoC alone. The cost-effectiveness of SGLT2i vs SoC is presented in B.3.8.3 and is consistent with Section 5.1.14 of the NICE Guide to Methods 2013(8). An economic comparison of empagliflozin vs dapagliflozin is not presented as it will not support prescriber choice or individualisation of care.

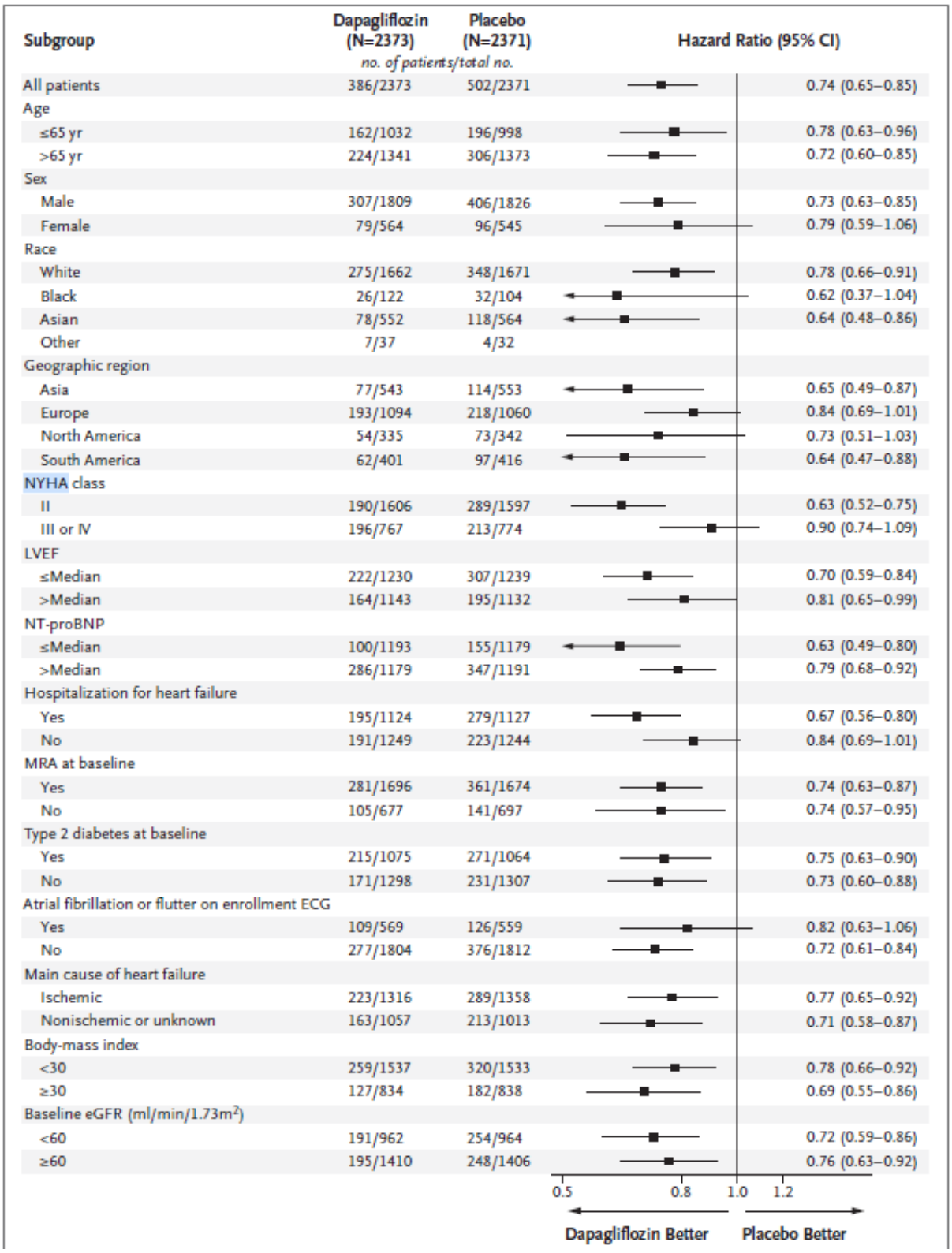
Table 28. Identified treatment effect modifiers

Endpoint	Identified treatment effect modifiers included in the adjusted analyses (categories matching on)
Time to first event of adjudicated CV death or adjudicated HHF	Race (White vs non-White) NYHA class (II vs III/IV) NT-pro-BNP (≤ median vs > median*) LVEF (≤ median vs > median*) eGFR (< 60, ≥ 60 ml/min/1.73 m^2)
Key: CV, cardiovascular; eGFR, estimated glomerular filtration rate; HHF, hospitalisation for heart failure; LVEF, left ventricular ejection fraction; N/A, not applicable; NT-pro-BNP, N-terminal prohormone brain natriuretic peptide; NYHA, New York Heart Association; vs, versus; 1, EMPEROR-Reduced clinical study report; 2, EMPEROR-Reduced exploratory analyses performed by BresMed; 3, DAPA-HF publication Note: * median based on median reported for comparator trial.	

Table 29. Weighting assigned to patients in EMPEROR-Reduced to be comparable to patients in DAPA-HF

Matching variable	Summary of characteristics			Weighting
	EMPEROR-REDUCED ITT	DAPA-HF ITT	Prognosis in EMPEROR-Reduced versus DAPA-HF	
Overall	NA	NA	NA	NA
Race				
White	70%	70%	Same	No change
Not White	30%	30%		No change
NYHA				
II	75%	68%	Less severe	[redacted]
III/IV	25%	32%		[redacted]
LVEF				
≤ median (32)	79%	50%	More severe	[redacted]
> median (32)	21%	50%		[redacted]
NT-pro-BNP				
≤ median (1437)	36%	50%	More severe	[redacted]
> median (1437)	64%	50%		[redacted]
eGFR				
< 60	48%	41%	More severe	[redacted]
≥ 60	52%	59%		[redacted]

Figure 18. Subgroup analyses for the primary outcome (a composite of hospitalisation for heart failure, an urgent visit resulting in intravenous therapy for heart failure, or death from cardiovascular causes) in DAPA-HF



B.2.9 Adverse events

Median exposure to study medication was approximately 14 months in both treatment groups, with 61% of patients treated for at least 1 year. Safety was assessed descriptively based on adverse events (AEs), adverse events of special interest (AESIs), and specific AEs.

A similar overall proportion of patients in the empagliflozin and placebo groups reported at least one AE, most of which were of mild or moderate intensity (Table 30). Proportions of patients experiencing severe AEs and AEs leading to premature discontinuation of study medication were also similar between the two groups (Table 31).

Table 30. Overall summary of AEs in the TS

Category of AEs	Placebo, N (%)	Empagliflozin 10 mg, N (%)
Number of patients in the TS, N (%)	1863 (100.0)	1863 (100.0)
Patients with any AEs	[redacted]	[redacted]
Mild	[redacted]	[redacted]
Moderate	[redacted]	[redacted]
Severe	[redacted]	[redacted]
Investigator-defined drug-related AEs	[redacted]	[redacted]
AEs leading to discontinuation of study medication	[redacted]	[redacted]
Serious AEs	[redacted]	[redacted]
Serious AEs		
Resulting in death	[redacted]	[redacted]
Life threatening	[redacted]	[redacted]
Persistent or significant disability/incapacity	[redacted]	[redacted]
Requires or prolongs hospitalisation	[redacted]	[redacted]
Congenital anomaly or birth defect	[redacted]	[redacted]
Other medically important serious event¶	[redacted]	[redacted]

Source: EMPEROR-Reduced CSR, Table 15.3.1.1 (135)
Abbreviations: AE, adverse event; TS, treated set.
Note: Percentages calculated using total number of patients per treatment as the denominator. A patient may be counted in more than one seriousness criterion.
¶Other medically important serious event was defined as any important medical event (when based upon appropriate medical judgment) which might jeopardise the patient and might require medical or surgical intervention to prevent one of the other serious outcomes included in the definition of serious adverse events shown in the table above. Examples of such events could be intensive treatment in an emergency room or at home for allergic bronchospasm, blood dyscrasias, or convulsions that do not result in hospitalisation or development of dependency or abuse.

The overall frequency of serious AEs (SAEs) was lower in the empagliflozin group than in the placebo group, consistent with the efficacy analyses of all-cause hospitalisations (Table 31). The most frequent SAEs were cardiac disorders, pneumonia and acute kidney injury. All other SAEs were reported in less than 3.0% of patients per treatment group.

Table 31. Serious AEs with frequency >1% -exposure adjusted, in the TS

MedDRA SoC MedDRA PT	Placebo, N (%)	Empagliflozin 10mg, N (%)
Number of patients	1863 (100%)	1863 (100%)
Total with SAEs	[redacted]	[redacted]
Cardiac disorders	[redacted]	[redacted]
Cardiac failure	[redacted]	[redacted]
Ventricular tachycardia	[redacted]	[redacted]
Atrial fibrillation	[redacted]	[redacted]
Cardiac failure congestive	[redacted]	[redacted]
Cardiac failure chronic	[redacted]	[redacted]
Cardiac failure acute	[redacted]	[redacted]
Acute myocardial infarction	[redacted]	[redacted]
Infections and infestations	[redacted]	[redacted]
Pneumonia	[redacted]	[redacted]
Renal and urinary disorders	[redacted]	[redacted]
Acute kidney injury	[redacted]	[redacted]
Renal impairment	[redacted]	[redacted]
Nervous system disorders	[redacted]	[redacted]
Ischaemic stroke	[redacted]	[redacted]

MedDRA SoC MedDRA PT	Placebo, N (%)	Empagliflozin 10mg, N (%)
General disorders & administration site conditions	[redacted]	[redacted]
Death	[redacted]	[redacted]
With investigator-defined drug-related AEs	[redacted]	[redacted]

Source: EMPEROR-Reduced CSR Table 12.2:1 (135)

Abbreviations: AE, adverse event; MedDRA, Medical dictionary for regulatory activities; SoC, system organ class; MedDRA PT, Medical dictionary for regulatory activities preferred term; SAE, serious adverse event; TS, treated set.

Adverse events of special interest (AESIs) were pre-specified in the protocol as hepatic injury, decreased renal function, ketoacidosis, and AEs leading to lower limb amputation. Overall frequencies of AESIs were comparable in the empagliflozin and placebo groups (Table 32). No ketoacidosis events were reported in either group.

Specific AEs were defined as urinary and genital tract infections, volume depletion and hypotension, confirmed hypoglycaemic events, bone fractures and urinary tract malignancies. As known for the drug class, uncomplicated genital tract infections occurred more often with empagliflozin than with placebo, while complicated genital infections or those leading to treatment discontinuation had similar frequency in both groups. There was a numerical but not clinically meaningful increase in volume depletion and hypotension with empagliflozin relative to placebo, including events that were reported as SAEs or that led to treatment discontinuation. No increase in confirmed hypoglycaemic events was detected for patients with or without T2DM, and no severe hypoglycaemic events were reported in patients without T2DM. The frequencies of the remaining types of specific AEs were similar between the groups (Table 32).

Table 32. Summary of AESIs and specific AEs, TS

Category of AESIs and specific AEs	Placebo, N (%)	Empagliflozin 10 mg, N (%)
Number of patients	1863 (100.0)	1863 (100.0)
AESIs		
Acute renal failure	[redacted]	[redacted]
Serious	[redacted]	[redacted]
Leading to discontinuation	[redacted]	[redacted]
Hepatic injury	[redacted]	[redacted]
Serious	[redacted]	[redacted]
Leading to discontinuation	[redacted]	[redacted]
Up to 30 days after treatment discontinuation	[redacted]	[redacted]
Ketoacidosis	[redacted]	[redacted]
AEs leading to LLA up to trial completion (investigator-defined)	[redacted]	[redacted]
Specific AEs		
Urinary tract infection	[redacted]	[redacted]
Complicated	[redacted]	[redacted]
Leading to discontinuation	[redacted]	[redacted]
Genital infection	[redacted]	[redacted]
Complicated	[redacted]	[redacted]
Leading to discontinuation	[redacted]	[redacted]
Volume depletion	[redacted]	[redacted]
Hypotension	[redacted]	[redacted]
Serious	[redacted]	[redacted]
Leading to discontinuation	[redacted]	[redacted]
Symptomatic hypotension (investigator-defined)	[redacted]	[redacted]
Confirmed hypoglycaemic events*	[redacted]	[redacted]
In patients with T2DM¶	[redacted]	[redacted]
In patients with pre-diabetes¶	[redacted]	[redacted]

Category of AESIs and specific AEs	Placebo, N (%)	Empagliflozin 10 mg, N (%)
In patients without diabetes or pre-diabetes¶	[redacted]	[redacted]
Bone fracture	[redacted]	[redacted]
Serious	[redacted]	[redacted]
Leading to discontinuation	[redacted]	[redacted]
Up to trial completion	[redacted]	[redacted]
Urinary tract malignancy up to trial completion	[redacted]	[redacted]

Source: Empagliflozin CSR Table 12.1.3:1 (135)

Abbreviations: AE, adverse event; AESI, adverse event of special interest; LLA, lower limb amputation; T2DM, type 2 diabetes mellitus; TS, treated set

*Hypoglycaemic AEs with a plasma glucose value of ≤70 mg/dL or where assistance was required

¶Patients with events/patients in subgroup (%)

B.2.10 Ongoing studies

There are no ongoing studies of empagliflozin relevant for this appraisal.

B.2.11 Innovation

As stated in B.1.3.1.2, HF affects just under 1 million people in the UK, of which nearly two-thirds are estimated to have HFrEF (42, 102). Although established treatments for HFrEF are associated with improved outcomes, the mortality and hospital admission rates of HFrEF patients remain high, with many patients not receiving recommended doses of treatment (57, 102, 153). ACE inhibitors or ARBs in combination with BBs are the cornerstone of the current management of HFrEF in the NHS, but optimal outcomes require gradual up-titration until the recommended dose is reached (34, 154, 155). In clinical practice, only a minority of patients achieve the recommended doses. This is often due to limiting comorbidities such as renal impairment, hypotension or asthma but could also be partly attributable to the length of time required to up-titrate with patients less likely to adhere to a lengthy treatment process without benefiting from immediate clinical improvements (61, 62). Patients on suboptimal ACEi/ARB or BB doses have been shown to have greater symptom burden and a significantly higher risk of death or HHF compared to the minority of patients on target doses (61). Sacubitril valsartan or ivabradine are recommended second line options for symptomatic patients (156, 157). Their uptake in clinical practice has however been limited by strict eligibility criteria (LVEF ≤ 35%, NYHA class II-IV and for ivabradine,

resting heart rate ≥75 bpm) and contraindications (hypotension, renal impairment and/or hyperkalemia) (158-160).

A critical challenge in the management of HFrEF therefore pertains to the large proportion of patients who do not receive guideline-directed doses of recommended treatments or are ineligible for them. This is typically an elderly population with pre-existing chronic kidney disease, diabetes and hypertension (10, 41), all of which are known risk factors for significantly worse outcomes and are a contributing factor to many patients being ineligible or unable to receive the recommended doses of HF therapy (113, 153, 161). A high unmet need for an effective treatment that improves HFrEF outcomes, symptoms and QoL among high-risk patients with CRM comorbidities therefore remains.

As an add-on to the SoC, empagliflozin offers a step change in the management of HFrEF within the NHS:

- It significantly reduces the risk of CV death or HHF while significantly improving renal outcomes and QoL in a population with broad spectrum of severity of HFrEF regardless of age, gender, use of neprilysin inhibitor, presence or absence of diabetes (113) or chronic kidney disease (111).
- As a fixed dose, once-per-day, orally administered medication, empagliflozin is simple for physicians to initiate and for patients to adhere to, saving NHS professionals' time that would otherwise be spent on dose titration or on training patients to self-administer.
- Substantial reduction in HHF seen with empagliflozin combined with simplicity of initiation suggests that its adoption in primary care could support efficiency improvements in the allocation of NHS resources by releasing capacity in secondary care. A recently published report by NICE on implementation of NG106 noted that patients with HF often have comorbid diabetes and CKD that require visits to additional specialist clinics (162). SGLT2 inhibitors like empagliflozin offer an opportunity to promote a more holistic approach to treatment of adults with T2DM (162). Empagliflozin is already indicated in T2DM (1), and with a marketing authorisation in HFrEF expected in August 2021, it could support this objective.

Having multiple SGLT2is recommended for use by NICE, similar to T2DM, has several benefits:

- To support continuation of care (i.e., no need to switch T2DM patients already managed with empagliflozin if they develop comorbid HFrEF).
- To overcome potential practical challenges, such as supply chain issues, should they occur.
- The availability of additional SGLT2is may help to bridge in inequalities seen between those with better or worse access to care (Section B.1.4).
- To allow patient and clinician choice.
- Tailor treatment for patients who are severely renally impaired. Unlike DAPA-HF, EMPEROR-Reduced permitted the inclusion of patients with an eGFR as low as 20mL/min/1.73m^2.
- Management of specific AE's. For example, unlike dapagliflozin, no cases of ketoacidosis were observed in the EMPEROR-Reduced trial; this assumption is further supported by the EMPA-REG OUTCOME trial in patients with T2DM and established CV disease where very few cases were observed and there was no imbalance between treatment groups (163).

B.2.12 Interpretation of clinical effectiveness and safety evidence

In the EMPEROR-Reduced trial, treatment with empagliflozin 10 mg once daily as an add-on to SoC in patients with HFrEF (LVEF≤40%) demonstrated superiority compared to placebo for the primary endpoint, time to the first occurrence of adjudicated CV death or adjudicated HHF. The superiority over placebo was also demonstrated for key secondary endpoints, occurrence of adjudicated HHF (first and recurrent) and eGFR (CKD-EPI)^{cr} slope of change from baseline.

Treatment with empagliflozin as an add-on to SoC leads to a clinically and statistically significant reduction in risk of CV death or HHF by 25% compared with placebo added to SoC. During the trial period, the number of patients who needed to be treated with empagliflozin to prevent one primary event was 19 (95% CI, 13 - 37). The treatment effect of empagliflozin became apparent shortly after randomisation and was

maintained throughout the trial. The results of all sensitivity analyses were consistent with the results of the primary analysis (i.e. the HR was numerically similar). The results were also consistent across the pre-defined subgroups stratified by baseline diabetes status, baseline eGFR, age and geographical region, as indicated by the point estimate HR for time to CV death or HHF for each subgroup being below the no-effect value of 1.

The risk of all-cause hospitalisation was reduced with empagliflozin compared to placebo, both for the first occurrence (by [redacted] and for recurrent events (by [redacted] Furthermore, fewer patients receiving empagliflozin were reported with all-cause or CV mortality, although the treatment effect was not significant on a nominal level. Most of the deaths in the trial were due to CV causes and were classified as sudden cardiac death or HF death, as expected in this population.

The decline in renal function, evaluated based on change in eGFR slope from baseline, was significantly slower in the empagliflozin group, with an estimated difference in slope of about 1.7 mL/min/1.73 m^2 per year vs. placebo. Consistent with the attenuated deterioration of renal function, the risk of serious renal outcomes (chronic dialysis, renal transplant, or sustained reduction in eGFR) was halved in the empagliflozin group relative to placebo, demonstrating that empagliflozin has clinically meaningful nephroprotective as well as cardioprotective effects. The findings from the EMPEROR-Reduced study therefore have important clinical implications for the holistic treatment of indications comprising the “cardiorenal syndrome”.

A higher proportion of patients in the empagliflozin group than in the placebo group showed a clinically meaningful improvement in KCCQ-CSS after 52 weeks of treatment of at least 5 points from baseline. Consistently, a lower proportion of patients in the empagliflozin group than in the placebo group showed deterioration. The favourable effect of empagliflozin was driven by all domains of the KCCQ-CSS, including symptom frequency, symptom burden, and physical limitations. Supportive analyses of KCCQ-OSS and KCCQ-TSS were consistent with these findings.

Overall, empagliflozin was well tolerated in HFrEF patients with or without T2DM. Adverse events reported in the trial were consistent with the known safety profile of empagliflozin. As expected for the SGLT2 drug class, uncomplicated genital infections

were more common in the empagliflozin group. The frequency of hypoglycaemia, lower limb amputation, and bone fracture did not differ between the two groups, even though these AEs have been associated with the use of SGLT2 inhibitors in trials with T2DM patients (164). Overall, the proportion of patients experiencing SAEs was lower in the empagliflozin than in the placebo group, consistent with the efficacy analyses of all-cause hospitalisations. Safety concerns that have been seen with other drugs for HF (e.g., hypotension, volume depletion, renal dysfunction, bradycardia, and hyperkalemia) were not evident with empagliflozin in EMPEROR-Reduced.

In addition to direct evidence, the effect of SGLT2 inhibition by empagliflozin or dapagliflozin on CV mortality and all-cause mortality in HFrEF was investigated in a meta-analysis of their RCTs (EMPEROR-Reduced and DAPA-HF, respectively) since neither trial was sufficiently powered to evaluate these endpoints (119). The meta-analysis estimated that SGLT2 inhibition was associated with a 13% reduction in all-cause death (pooled HR 0·87, 95%CI, 0.77–0.98; p=0·018) and 14% reduction in CV death (pooled HR 0·86, 95%CI, 0·76–0·98; p=0·027). The risk of a composite renal endpoint was also significantly reduced with SGLT2 inhibitors (0·62, 0·43–0·90; p=0·013). Since EMPEROR-Reduced was enriched for patients with more severe but stable HFrEF, it provided evidence that benefits of SGLT2 inhibition extend to patients with severe left ventricular dysfunction.

In conclusion, data presented in this section demonstrate that empagliflozin 10mg is associated with a clinically meaningful reduction in risk of CV death or HHF and a slower progressive decline of renal function in patients with HFrEF, regardless of the presence or absence of diabetes. The data therefore supports addition of empagliflozin to the guideline-directed medical therapy for this patient population.

B.3 Cost-effectiveness

- A Markov model with health states defined by KCCQ-CSS quartiles was developed to estimate the lifetime costs and outcomes of patients with HFrEF.
- In base case analysis, which evaluated the cost-effectiveness of empagliflozin as an add-on to SoC compared to SoC alone, empagliflozin was estimated to increase life years and quality adjusted life years by [redacted] and [redacted] per patient, respectively, and to reduce HHF by [redacted] events per 100 patient-years. The incremental cost-effectiveness ratio of [redacted] per QALY indicated that empagliflozin is highly cost-effective as an addition to the SoC in HFrEF.
- Deterministic sensitivity analyses demonstrated that the cost-effectiveness results were robust with respect to variation in individual model parameters. The treatment effect associated with HHF was identified as the most influential driver of model results. The ICER however remained below [redacted] (<£10k) per QALY across all parameter variations.
- Probabilistic sensitivity analyses indicated [redacted] and [redacted] probability of empagliflozin being cost-effective at the willingness-to-pay thresholds of £20,000 per QALY and £30,000 per QALY, respectively.
- Results of scenario analyses indicate that ICER is not significantly affected by structural assumptions including the number of inflection points, choice of parametric distribution for mortality or treatment discontinuation, utility age-adjustment, or cost of non-CV death, with all scenarios resulting in ICERs <£6,500 per QALY.
- Empagliflozin therefore represents a highly cost-effective use of NHS resources in the treatment of symptomatic patients with chronic HF and reduced ejection fraction.

B.3.1 Published cost-effectiveness studies

A SLR was conducted to identify existing economic evaluations for the treatment of HFrEF patients. Full details of the process and methods used are described in Appendix G.

In summary, a total of 44 cost-effectiveness studies were identified. Of these, nine were conducted in the UK and deemed relevant for this submission. A table summary of all included studies and a critical appraisal of the nine UK-based studies are summarised in Appendix G. Table 33 reports a summary of the model characteristics, patient population and results of the UK-based studies, two of which were NICE technology appraisals, three were Scottish Medicines Consortium (SMC) technology appraisals, one was an All Wales Medicines Strategy Group (AWMSG) technology appraisal and three were publications. All studies reported the type of model used in the analysis, and these varied from a Markov model (often reported as two state) to a discrete event simulation. Where reported, a monthly cycle length was commonly modelled over a lifetime horizon. The cost-effectiveness analysis and modelling approach relevant to the decision problem was aligned with a recent McEwan et al 2020 publication (165).

Table 33. Summary list of published UK cost-effectiveness studies

Study, Year	Summary of model	Patient population (average age in years)	QALYs (intervention, comparator)	Costs (currency) (intervention, comparator)	ICER (per QALY gained)
McEwan et al 2020 (165)	**Type of model:** Markov state-transition cohort model with health states stratified by baseline T2DM status. **Time horizon:** Lifetime **Cycle length:** Monthly	Patients from the DAPA- HF trial: Aged ≥18 years with NYHA functional class II to IV HF, with a left ventricular ejection fraction ≤40%, and were optimally treated with pharmacological and device therapy. Average age of 66 years for Intervention (Dapagliflozin + standard therapy) and 67 years for comparator (standard therapy)	Intervention (Dapagliflozin plus standard therapy): 4.61 Comparator (standard therapy): 4.13	Intervention (Dapagliflozin plus standard therapy): £16,408 Comparator (standard therapy): £13,628	£5,822 (Committee's preferred ICER was £7,264)
McMurray et al 2018 (166)	**Type of model:** Decision analytic model developed based on a series of regression models **Time horizon:** Lifetime **Cycle length:**	Patients based on the characteristics in the PARADIGM-HF trial, with mean age of 64 years	Intervention (Sacubitril valsartan): 5.58 Comparator (ACEi): 5.06	Intervention (Sacubitril valsartan): £23,720 Comparator (ACEi): £14,814	£17,134

Study, Year	Summary of model	Patient population (average age in years)	QALYs (intervention, comparator)	Costs (currency) (intervention, comparator)	ICER (per QALY gained)
	1 month (with half-cycle correction)				
Lee et 2014 (167)	**Type of model:** Discrete event simulation **Time horizon:** Lifetime **Cycle length:** Not reported	Simulated patient population derived from that enrolled in EMPHASIS-HF trial (all patients had NYHA class II, with a mean age of 69 years, a mean LVEF of 26% and 78% patients were men)	Intervention (Eplerenone plus standard care): 6.19 Comparator (standard care alone): 4.98	Intervention (Eplerenone plus standard care): £18,559 Comparator (standard care alone): £14,275	£3,520
NICE [Sacubitril valsartan], 2015 (168)	**Type of model:** Two state Markov economic model with health states defined as alive and dead **Time horizon:** Lifetime **Cycle length:** 1 month (with half-cycle correction)	Individual patient-level data from the PARADIGM-HF trial with a mean age of 64 years	Intervention (Sacubitril valsartan in combination with standard care): 4.87 Comparator (ACEi in combination with standard care): 4.46	Intervention (Sacubitril valsartan in combination with standard care): £20,734 Comparator (ACEi in combination with standard care): £13,286	£18,187
NICE [Ivabradine], 2012 (169)	**Type of model:** Two state Markov economic model with health states defined as alive and dead **Time horizon:** Lifetime **Cycle length:**	Patient-level data from the SHIFT trial with a mean age of 61 years	Incremental QALYs (ivabradine plus standard care vs. standard care alone): 0.28	Incremental costs (ivabradine plus standard care vs. standard care alone): £2,376	£8,498

Study, Year	Summary of model	Patient population (average age in years)	QALYs (intervention, comparator)	Costs (currency) (intervention, comparator)	ICER (per QALY gained)
	Monthly (with half-cycle correction)				
SMC [Sacubitril valsartan], 2016 (170)	**Type of model:** Two state Markov economic model with health states defined as alive and dead **Time horizon:** Lifetime **Cycle length:** Not reported	Patients were modelled using the baseline characteristics of each patient from the PARADIGM-HF study with mean age of 64 years	Incremental QALYs (sacubitril valsartan plus standard care vs. enalapril plus standard care): 0.42	Incremental costs (sacubitril valsartan plus standard care vs. enalapril plus standard care): £7,685	£18,348
SMC [Eplerenone], 2012 (171)	**Type of model:** Decision analytic model developed based on a series of regression models **Time horizon:** Lifetime **Cycle length:** Not reported	Patients similar to that of the EMPHASIS-HF study with NYHA class II HF and LVEF≤30% with mean age of 69 years	Incremental QALYs (sacubitril valsartan plus standard care vs. enalapril plus standard care): 1.21	Incremental costs (sacubitril valsartan plus standard care vs. enalapril plus standard care): £3,140	£3,822
SMC [Ivabradine], 2012 (172)	**Type of model:** Markov model **Time horizon:** Not reported **Cycle length:** Not reported	Patients with chronic HF NYHA class II-IV in sinus rhythm and heart rate ≥75 bpm with a mean age of 60 years	Incremental QALYs (ivabradine plus standard care vs. standard care alone): 0.31	Incremental costs (ivabradine plus standard care vs. standard care alone): £1,875	£6,002

Study, Year	Summary of model	Patient population (average age in years)	QALYs (intervention, comparator)	Costs (currency) (intervention, comparator)	ICER (per QALY gained)
AWMSG [Eplerenone], 2012 (173)	**Type of model:** Discrete event simulation model **Time horizon:** Lifetime **Cycle length:** Not reported	Patients with chronic systolic HF with NYHA class II symptoms and LVEF≤ 30%, in line with the patient population in the EMPHASIS-HF trial with mean age of 69 years	Intervention (eplerenone with standard care): 6.19 Comparator (standard optimal therapy): 4.98	Intervention (eplerenone in combination with standard care): £14,184 Comparator (standard optimal therapy): £9,882	£3,534

Abbreviations: AWMSG, All Wales Medicines Strategy Group; HF, heart failure; LVEF, left ventricular ejection fraction; NYHA, New York Heart Association; QALYs, quality-adjusted life years; SMC, Scottish Medicines Consortium; T2DM, type 2 diabetes mellitus; ICER, incremental cost-effectiveness ratio

B.3.2 Economic analysis

The SLR of cost-effectiveness studies, described in Appendix G, identified 44 unique economic evaluations in HFrEF. None of the published economic models were available for the evaluation of the cost-effectiveness of empagliflozin versus placebo, and of SGLT2 inhibitors versus placebo, in adults with chronic HFrEF. Therefore, a *de novo* model was developed using Microsoft Excel® (Office 365, version 2008) with Visual Basic for Applications (VBA) functionality.

The published models for HF treatments were considered during the conceptualisation of the model for empagliflozin. In particular, the economic model for dapagliflozin in HFrEF submitted to NICE as part of TA679 was found to adequately reflect the variation in risk with disease severity through the use of time-updated KCCQ-TSS covariates in risk equations for all-cause mortality, CV mortality, HHF (6). The NICE evaluation committee for TA679 concluded that the KCCQ tool is a reasonable way to classify disease severity and is appropriate for decision making (6). A similar approach was therefore adopted for modelling cost-effectiveness of empagliflozin in HFrEF, that is, with KCCQ-CSS rather than KCCQ-TSS-defined health states as explained in more detail in section B.3.2.2.

B.3.2.1 Patient population

The patient population considered in the cost-effectiveness analysis is adults with symptomatic chronic heart failure with reduced ejection fraction in accordance with the anticipated marketing authorisation of empagliflozin and the decision problem considered in this submission. Empagliflozin is indicated in adults for the treatment of symptomatic chronic heart failure with reduced ejection fraction. The population is also reflective of the ITT population of the EMPEROR-Reduced trial (109).

In accordance with the trial inclusion criteria, the modelled cohort comprised adults with chronic HF with LVEF ≤ 40% and NYHA class II-IV. The KCCQ-CSS distribution of patients in the ITT population at baseline across quartiles was used to inform the initial distribution of patients across alive health states at the start of the model and influenced the rates of all-cause death, CV death, and HHF (Table 34). The modelled

cohorts in the empagliflozin + SoC and placebo + SoC arms were assigned the same baseline characteristics.

Table 34. Mean patient characteristics of the modelled cohort at model entry based on EMPEROR-Reduced trial, ITT population

Baseline characteristic	ITT population	SE
Demographics		
Age (years)	66.84	0.18
Age (≥65 years)	62%	0.01
Sex: Male	76%	0.01
Region		
Asia	13.2%	0.01
Europe	36.3%	0.01
Latin America	34.5%	0.01
North America	11.4%	0.01
Other	4.6%	0.00
KCCQ-CSS		
KCCQ-CSS 0 to <55 (Quartile 1)	[redacted]	[redacted]
KCCQ-CSS: 55 to <75 (Quartile 2)	[redacted]	[redacted]
KCCQ-CSS: 75 to <90 (Quartile 4)	[redacted]	[redacted]
KCCQ-CSS: 90 to 100 (Quartile 4)	[redacted]	[redacted]
NYHA class		
Baseline NYHA II	75.1%	0.01
Baseline NYHA III	24.4%	0.01
Baseline NYHA IV	0.5%	0.00
Treatment use at baseline		
ACEi	45.4%	0.01
ARB	24.3%	0.01
ARNi	19.5%	0.01
MRA	71.3%	0.01
BB	94.7%	0.00
Loop or high ceiling diuretics (furosemide)	84.5%	0.00
Medical history		
Ischaemic cause of HF	51.7%	0.01

Abbreviations: ACEi, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; ARNi, angiotensin receptor neprilysin inhibitor; BB, beta-blocker; CSS, clinical summary score; HF, heart failure; ITT, intent to treat; IVA, ivabradine; KCCQ, Kansas City Cardiomyopathy Questionnaire; MRA, mineralocorticoid receptor antagonist; NYHA, New York Heart Association; SE, standard error.

B.3.2.2 Model structure

The model uses a Markov cohort state-transition approach and describes the clinical course of HFrEF using five discrete health states defined by quartiles of the baseline distribution of KCCQ-CSS in the combined empagliflozin and placebo treatment groups (KCCQ-CSS quartiles 1 to 4 corresponding to KCCQ-CSS scores of 0 to [redacted] to [redacted], and [redacted] to 100, respectively, with higher score corresponding to a better health status), and death, with health state-specific costs and utilities (Figure 19). The use of quarters vs tertiles was also explored for categorising KCCQ-TSS. Quartiles were found to provide a better fit to the observed data than tertiles while still retaining adequate patient numbers in each subgroup to permit statistically robust analysis and providing sufficient granularity in predicting patient outcomes. Evenly spaced quarters were also rejected (i.e. 0-25, 26-50, 51-75, 76-100) as they did not contain adequate patient numbers in each group for a robust analysis. Similarly, health states defined by KCCQ-CSS tertiles of the baseline distribution was explored; however the analysis of transition probabilities showed less differentiation between the treatment groups and over time, suggesting loss of sensitivity to differences.

The patient cohort entered the model according to the baseline distribution of KCCQ-CSS quartiles. From this state, patients could transition to a higher (i.e., regress/lower disease burden) or lower (i.e., progress/higher disease burden) KCCQ-CSS quartile, remain in the same state, or die. In each of the states, patients could experience an AE or HHF, or a composite renal outcome. Transitions between the health states occurred in one month cycles, and half-cycle correction was applied.

KCCQ score is an established disease-specific measure of health status derived from a 23-item self-administered questionnaire that quantifies a patient's perception of their health status (174, 175). The KCCQ score has been shown to be valid, reliable, and sensitive to clinical changes, with low KCCQ score being an independent predictor of poor prognosis in HF (176-178). KCCQ-CSS, unlike KCCQ-TSS, was an exploratory endpoint in EMPEROR-Reduced. Stratification of patients by baseline KCCQ-CSS showed that the risk of CV death or HHF was higher in patients with lower baseline KCCQ-CSS ([redacted], and [redacted] per 100 patient years at risk for KCCQ-CSS score <[redacted], respectively) (120). This prognostic correlation enabled the impact of disease severity

to be captured in the KCCQ-CSS health state utilities and risk of events, thus allowing more accurate modelling of the clinical course of HF compared with the two state Markov model in TA388 (168).

Health states defined by KCCQ-CSS rather than NYHA class or KCCQ-TSS represent an improvement of the cost-effectiveness models in TA267 and TA679 (6, 169) because:

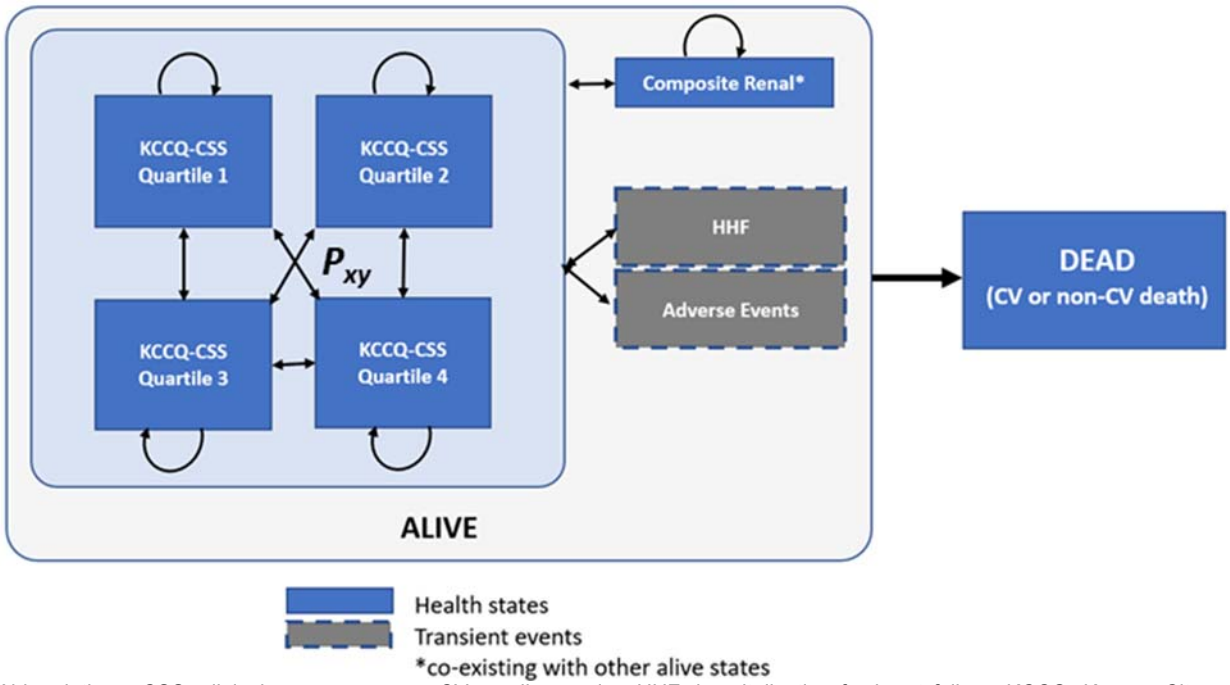
- As a patient-reported outcome, KCCQ-CSS is a more accurate and reproducible measure of HF severity than the physician-reported NYHA class (32);
- Compared to KCCQ-TSS, KCCQ-CSS is broader in scope as it encompasses TSS and physical limitations (174).

The model captured the occurrence of first and subsequent HHF and treatment-related AEs as transient events. Transition to the death state was modelled using parametric survival equations for CV mortality and all-cause mortality. All model equations were derived using the EMPEROR-Reduced trial data with KCCQ-CSS health states as time-varying predictors.

At the end of each cycle, patients transitioned from the alive health states to death based on the estimated all-cause death rate. The CV death equation was used to estimate the proportion of patients who die from CV causes. The difference between the all-cause death rate and the CV death rate represented the non-CV death rate. Patients could discontinue treatment with empagliflozin at any cycle. After discontinuation, patients received SoC treatment until death or the end of the model time horizon. Patients who discontinued treatment with empagliflozin experienced thereafter the same event rates and health state transition probabilities as patients receiving placebo. The transition probability matrix for transitions between KCCQ-CSS quartiles was then applied to the remaining patients in the alive health states to calculate the health state distribution in the next cycle. Monthly transition probabilities were derived using longitudinal measurements of health status defined by KCCQ-CSS in the trial, as described in section B.3.3. Within each alive health state, patients could experience an adverse renal outcome (chronic dialysis, renal transplant or sustained

reduction in eGFR) according to the event rates observed in the EMPEROR-Reduced trial (Section B.2.6.2.3).

Figure 19. Model schematic



Abbreviations: CSS, clinical summary score; CV, cardiovascular; HHF, hospitalisation for heart failure; KCCQ, Kansas City Cardiomyopathy Questionnaire. Q1, Quartile 1: 0 to <55.2; Q2, Quartile 2: 55.2 to <75; Q3, Quartile 3: 75 to <89.6; Q4, Quartile 4: 89.6 to 100.

Costs included direct medical costs for treatment acquisition, clinical event management and disease management. Utilities were accrued based on time spent in each KCCQ-CSS quartile, adjusted for disutilities associated with HHF and AEs. The model tabulates the cumulative number of clinical events experienced by the cohort, event rates per 100 person years (PY), life years (LY), QALYs, costs, and ICER. A 3.5% annual discount rate was applied to costs and health outcomes

A Markov multi-state model structure based on disease severity was considered the most appropriate because it allowed explicit modelling of the relationship between disease progression and clinical outcomes through the specification of different rates of HHF and CV death depending on the KCCQ-CSS quartile, and enabled inclusion of short-term as well as long-term health benefits in the rate of health state transitions. This structure addresses the concerns surrounding the two state Markov models of chronic HF previously submitted to NICE (168, 169) and follows the well-received

modelling approach in TA679 (6). It was also considered simpler and more efficient than a patient-level simulation approach (used in NICE TA388 (168) requiring less computational time while still adequately capturing heterogeneity across patients with HFrEF through a tractable number of mutually exclusive and exhaustive health states whose occupancy is determined by patient characteristics related to individual's KCCQ-CSS state. The model presented here also captures empagliflozin's capacity to slow the progression of renal impairment in accordance with the outcomes specified in the Final Scope.

The chosen model structure is aligned with the clinical care pathway described in section B.1.3.1.1 and reflects the anticipated early positioning of empagliflozin as an add-on to the current first line treatments for HFrEF recommended in the NICE treatment guideline for heart failure (ACEi/ARBs + BBs ± MRA, with ARB prescribed for patients intolerant to ACEis) (23). The main features of the cost-effectiveness model are summarised and compared to those of previous NICE technology appraisals in HFrEF in Table 35.

Table 35. Features of the current and previous economic analysis

	Previous appraisals			Current appraisal	
Factor	**TA267 (Ivabradine)**	**TA388 (Sacubitril valsartan)**	**TA679 (Dapagliflozin)**	**Chosen values**	**Justification**
Model structure	A 2-state Markov cohort model (alive and dead). Within the alive state, patients were further sub-divided into 4 NYHA states	Patient-level simulation in the base-case; secondary analyses used a 2-state Markov cohort model (alive and dead)	A Markov model with 5 states, including death, using KCCQ-TSS quartiles to capture disease severity and progression	A Markov model with 5 states, including death, using KCCQ-CSS quartiles to capture disease severity and progression	A time-varying covariate by KCCQ-CSS quartile in risk/survival /utility equations allows modelling of the relationship between disease severity and outcomes (HHF, CV death, all-cause death)
Comparators	Standard treatment without ivabradine.	ACEi in combination with standard care. ARB in combination with standard care (for people in whom an ACEi is unsuitable). Standard care includes treatment with a beta-blocker and an aldosterone antagonist.	For the treatment of HFrEF patients on ACEi or ARB, in combination with beta-blocker, ±MRA, the comparators are: • Sacubitril valsartan • Placebo For the treatment of HFrEF patients on sacubitril valsartan, in combination with beta-blocker, ±MRA, the comparator is: • Placebo	Standard care without empagliflozin. Standard care is defined as: • ACE inhibitors in combination with beta-blockers, and/or mineralocorticoid receptor antagonists • ARBs in combination with beta-blockers, and/or mineralocorticoid receptor antagonists Sacubitril valsartan in combination with beta-blockers, and/or	See Table 1.

	Previous appraisals			Current appraisal	
Factor	**TA267 (Ivabradine)**	**TA388 (Sacubitril valsartan)**	**TA679 (Dapagliflozin)**	**Chosen values**	**Justification**
				mineralocorticoid receptor antagonists	
Time horizon	Lifetime	Lifetime	Lifetime	Lifetime	HF is a chronic disease, with costs and effects of treatment accumulating over lifetime
Treatment waning effect?	No	No	No	No	No evidence of treatment waning in EMPEROR-Reduced
Source of utilities	SHIFT trial	Baseline utilities from PARADIGM trial; rate of EQ-5D decline from Berg 2015 explored in scenario analyses (179)	DAPA-HF trial	EMEPEROR-Reduced trial	As per NICE reference case (144)
Source of costs	NHS and PSS price sources, and literature for other cost inputs	NHS and PSS price sources, and literature for other cost inputs	NHS and PSS price sources, and literature for other cost inputs	NHS and PSS price sources, and literature for other cost inputs	As per NICE reference case (144)
Perspective on health effects	Direct health effects	Direct health effects	Direct health effects	Direct health effects	As per NICE reference case (144)
Perspective on costs	NHS and PSS	NHS and PSS	NHS and PSS	NHS and PSS	As per NICE reference case (144)
Discounting	3.5%	3.5%	3.5%	3.5%	As per NICE reference case (144)

	Previous appraisals			Current appraisal	
Factor	**TA267 (Ivabradine)**	**TA388 (Sacubitril valsartan)**	**TA679 (Dapagliflozin)**	**Chosen values**	**Justification**
Cycle length	One month, with half-cycle correction	One month, with half-cycle correction	One month, with half-cycle correction	One month, with half-cycle correction	The shortest practical cycle length, given the frequency of trial data collection and a lifetime horizon

Abbreviations: CV, cardiovascular; HHF, hospitalisation for heart failure; HF, heart failure; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire – clinical summary score; KCCQ-TSS, Kansas City Cardiomyopathy Questionnaire – total symptom score; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; PSS, Personal and Social Services; QALY, quality-adjusted life year.

B.3.2.3 Intervention technology and comparators

The base-case analysis in this submission compares empagliflozin, as an add-on to SoC, in patients who are already treated with either;

- ACEi or ARB in combination with a BB, with or without MRA, or
- Sacubitril valsartan in combination with a BB, with or without MRA

to SoC alone. This comparison is in line with the proposed positioning of empagliflozin in section B.1.3.1.3 and the current NICE guideline for the management of chronic HF (23).

The cost-effectiveness analysis of empagliflozin+SoC vs sacubitril valsartan+SoC is not warranted since empagliflozin is not intended to replace but instead be used in addition to the sacubitril valsartan-based SoC. Clinical experts view SGLT2is and ARNi as two separate pillars of care that could be used concurrently to yield additional benefits (103, 180). The British Society for Heart Failure has recently appealed against sacubitril valsartan being considered as a replacement to dapagliflozin in TA679 (6).

Dapagliflozin, an SGLT2 inhibitor with clinical efficacy consistent with that of empagliflozin in heart failure, is not a relevant comparator as it does not represent the NHS SoC and has very low usage in HF patients, with market shares driven by its use in HF with comorbid T2DM (150) (Table 2). No head-to-head comparison of efficacy of empagliflozin vs dapagliflozin in HFrEF has been conducted, and attempts at indirect comparison have produced biased estimates of relative effectiveness due to differences in the severity of EMPEROR-Reduced and DAPA-HF trial populations that could not be adequately corrected using population adjustment methods (Section B.2.8). Since available evidence indicates clinical equivalence of empagliflozin and dapagliflozin in HFrEF with LVEF ≤ 40%, a scenario analysis was conducted to compare SGLT2 inhibitors in combination with SoC to standard SoC alone using treatment effect estimates from a pooled meta-analysis of DAPA-HF and EMPEROR-Reduced trials (119). This is consistent with the NICE Guide to Methods 2013(144) which states that "*if comparators form part of a class of treatments, and evidence is available to support their clinical equivalence, estimates of QALYs gained for the class*

as a whole can be presented" (Section 5.1.14) (8). The use of this larger data set enabled the uncertainty surrounding CV mortality and all-cause mortality efficacy estimates for SGLT2is to be reduced, given that neither EMPEROR-Reduced nor DAPA-HF were powered to detect statistical significance in these endpoints (109, 150).

B.3.3 Clinical parameters and variables

The clinical inputs for the economic model of empagliflozin were derived from analyses of patient-level data for the ITT population of EMPEROR-Reduced and consisted of the following:

- Transition matrices describing the probability of moving to each of the KCCQ-CSS quartile health states over time with empagliflozin + SoC treatment and SoC alone, given the current $P_{x,y}$ (Figure 19).
- Projected survival distributions for all-cause and CV mortality as a function of current health state and treatment implemented as parametric survival equations with treatment and time-varying health state indicators;
- Rate of HHF over time as a function of current health state and treatment implemented as a repeated measures Poisson regression with treatment and time-varying health state indicators;
- Change in utility associated with HHF and AEs derived from mixed-effects regression analyses relating the occurrence and timing of these events to changes in utilities;
- Treatment discontinuation with empagliflozin implemented as a parametric survival equation.

The choice of predictors in the risk equations was guided with the aim of preserving alignment between the observed and the predicted outcomes in the cohort Markov model. Following extensive baseline variable testing, the treatment and time-varying health state indicators were retained as predictors in the risk equations. Since the treatment effect of empagliflozin was found to be consistent across subgroups defined by age, gender, body-mass index, race, presence or absence of diabetes, baseline eGFR, and prior therapies (ARNi or MRA) to that observed in the ITT population

(Section B.2.7 and Appendix E), only the ITT population was considered in the economic analysis.

B.3.3.1 Baseline characteristics

The risk equations for all-cause mortality, CV mortality, HHF, and treatment discontinuation underlying the model predict the expected outcomes in a population where individuals are homogeneous in terms of the set of predictors in the equations. A cohort Markov structure is inherently designed for modelling homogeneous populations where a profile defined by the mean of the patient characteristics closely resembles the true study population (181, 182). The baseline characteristics of the modelled cohort, representative of the ITT EMPEROR-Reduced population, are shown in Table 34. While KCCQ-CSS is included in risk equations directly as a time-varying predictor, the association of other baseline variables with outcomes is captured indirectly, through their correlation with KCCQ-CSS. Thus, the Markov cohort is fully defined with regards to the average characteristics of the corresponding ITT population in EMPEROR-Reduced and the observed outcomes in this population are directly comparable with the predicted results. This is achieved at the expense of loss of flexibility, however, since the model can only be run in the population for which equations have been derived.

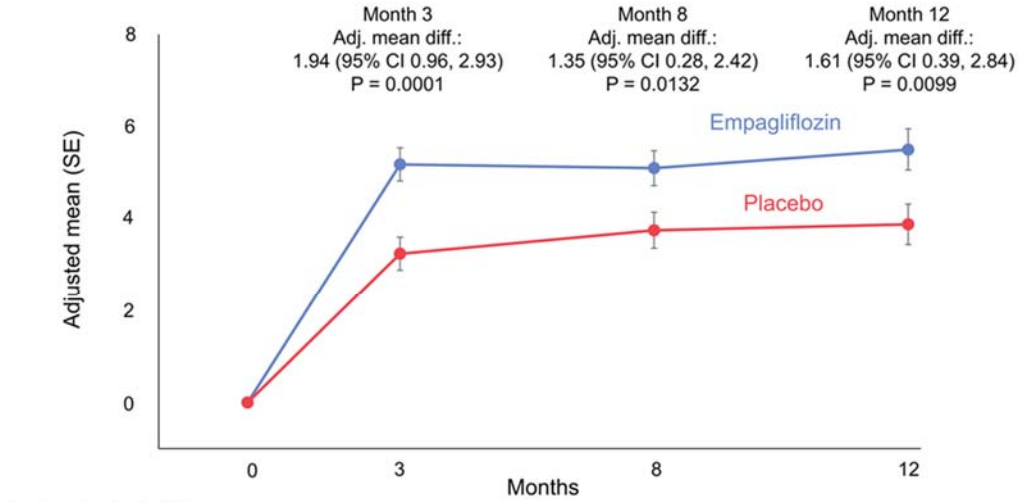
B.3.3.2 Health state transition probabilities

Treatment effect was found to be statistically significant with respect to change in KCCQ-CSS from baseline: the average KCCQ-CSS score increased by 5.7 points in the empagliflozin + SoC arm vs 4.3 points among patients on SoC alone by week 52 (Figure 20). In the ITT population, patients were most likely to remain at their current health status level in both arms; any changes in health status were more likely to occur in the first three months of treatment and tended to remain relatively stable thereafter. Improvement in levels tended to be more likely than declines in both treatment groups, with probabilities generally more favourable among patients receiving empagliflozin (Figure 20).

Treatment specific changes in health status were captured in the model through KCCQ-CSS quartile transition probabilities, with treatment specific transition matrices

derived from the analysis of KCCQ-CSS data collected in EMPEROR-Reduced at baseline and at weeks 12, 32, and 52. Analysis of transition probabilities consisted of deriving the proportion of ITT population in each KCCQ-CSS health state at a given time stratified by the previous health state based on longitudinal measurements of KCCQ-CSS health status in the trial with imputation by last observation carried forward strategy for missing visits while patients were still alive and followed. Missing measurements due to early end of follow-up were not imputed since death status past the end of follow-up was unknown and the distribution of the last known KCCQ-CSS health states for patients with early end of follow-up was similar to the distribution among observed/imputed data. The observed transition probabilities were found to vary over the three time periods (baseline-week 12, week 12–32 and week 32–52), revealing inflection points at week 12 and week 32 (Table 36). Therefore, three sets of period-specific probabilities were used in the model. Each of the three derived matrices was then converted to monthly transition probabilities by finding the m-root of the observed transition matrix for a longer period (e.g., 12 weeks, 20 weeks). This yielded three sets of monthly transition probabilities representing progression in the three periods used in the analysis (Table 36) (120). The model uses the monthly transition matrices from the last period (month 9+) to predict progression after the first year, assuming the probabilities remain constant in the long-term.

Figure 20. Effect of empagliflozin vs placebo on mean KCCQ-CSS over time



Analyzed patients (n)				
Placebo	1753	1732	1568	1218
Empagliflozin	1776	1755	1618	1239

Source: Butler et al 2021 *(120)*

Table 36. Monthly KCCQ-CSS transition matrix

KCCQ-CSS transitions [From, To]	Empagliflozin + SoC			Placebo + SoC		
	Months 1-3	Months 4-8	Months 9+	Months 1-3	Months 4-8	Months 9+
KCCQ [1,1]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [1,2]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [1,3]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [1,4]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [2,1]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [2,2]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [2,3]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [2,4]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [3,1]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [3,2]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [3,3]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [3,4]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [4,1]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [4,2]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [4,3]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [4,4]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: CSS, clinical summary score; KCCQ, Kansas City Cardiomyopathy Questionnaire; SoC, standard of care.

B.3.3.3 Mortality

B.3.3.3.1 CV mortality and all-cause mortality

A parametric survival analysis was conducted to allow for extrapolation of time to all-cause death and CV-related death as a function of treatment and time-varying KCCQ-CSS health states as measures of disease progression beyond the EMPEROR-Reduced trial duration. The analysis was conducted following recommendations of the NICE Decision Support Unit Technical Support Document 14 (183).

Selection of the best fitting parametric model consisted of fitting exponential, Weibull, Gompertz, log-logistic, log-normal, and the generalised gamma distributions to the observed data, including fitting of different models by treatment arm to explore and account for possible non-proportionality of effects. The goodness-of-fit was evaluated by graphical assessment of diagnostic plots, fit statistics [Akaike Information Criteria with correction for a finite sample size (AICc) and Bayesian Information Criteria (BIC)] and clinical plausibility of extrapolations. Time-varying indicators of the current health state were then introduced to a model based on the selected parametric distribution.

The joint Weibull model (i.e. the same parametric model fitted through both treatment arms) was selected as the base-case distribution for all-cause and CV death. For all-cause mortality, the choice was based on Weibull being the best fitting distribution ((although differences in AIC/BIC between fits were very small; (Table 37)) and yielded the most clinically plausible estimates of long-term survival (mean life expectancy of [redacted] months with placebo and [redacted] months with empagliflozin). To ensure the correct ordering of CV mortality and all-cause mortality predictions at all times, Weibull distribution was also chosen for modelling CV mortality although a comparable fit was obtained with the log-logistic distribution (Table 37). However, log-logistic distribution yielded improbably long mean survival predictions (mean survival of [redacted] months with placebo and [redacted] months with empagliflozin) compared to Weibull predictions which had greater clinical validity (mean survival of [redacted] months with placebo and [redacted] with empagliflozin) (Figure 21, **Error! Reference source not found.**).

Table 37. Goodness-of-fit statistics (AICc/BIC) for alternative parametric distributions, all-cause mortality, and CV mortality

Distribution	All-cause mortality		CV mortality	
	AICc	BIC	AICc	BIC
Exponential	[redacted]	[redacted]	[redacted]	[redacted]
Weibull	[redacted]	[redacted]	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]	[redacted]	[redacted]
Log-normal	[redacted]	[redacted]	[redacted]	[redacted]
Log-logistic	[redacted]	[redacted]	[redacted]	[redacted]
Generalised gamma	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: AICc, corrected Akaike Information Criterion; BIC, Bayesian Information Criterion.

Figure 21. Kaplan-Meier curves for all-cause mortality from EMPEROR-Reduced and extrapolated survival curves (Weibull)

[redacted]

Abbreviations: KM, Kaplan-Meier curve; SoC, standard of care.

Figure 22. Kaplan-Meier curves for CV mortality from EMPEROR-Reduced and extrapolated survival curves (Weibull)

[redacted]

Abbreviations: KM, Kaplan-Meier curve; SoC, standard of care.

Coefficients of the Weibull risk equations for all-cause and CV mortality with time-updated KCCQ-CSS health states are shown in Table 38.

Table 38. Parameterisation of survival equations for CV and all-cause death, ITT population of EMPEROR-Reduced trial, Weibull distribution (base-case)

Parameter	All-cause death			CV death		
	Coefficients	SE	p-value	Coefficients	SE	p-value
Shape	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Scale	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Treatment effect Empagliflozin 10 mg (Ref.: Placebo)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 55 to <75 (Quartile 2)*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 75 to <90 (Quartile 3)*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 90 to 100 (Quartile 4)*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: CV, cardiovascular; CSS, clinical summary score; KCCQ, Kansas City Cardiomyopathy Questionnaire; SE, standard error.

* vs KCCQ-CSS: 0 to 55 (Quartile 1)

Other parametric distributions for all-cause mortality and CV mortality had the following disadvantages compared to Weibull:

- The constant hazard assumed by the exponential distribution was not realistic for a progressive disease, where the risk of CV events and death is likely to increase over time
- The log-normal and log-logistic distributions showed improbably long predicted mean survival times of around over 30 years in both arms in a population that is mostly over 65 years old, and improbably long predicted mean time to CV death
- The Gompertz fit yielded short survival predictions (less than five years in both arms) and a sharply increasing hazard of CV death, which likely overestimated risk in this population
- The joint fits with generalised gamma distribution were more plausible and comparable to those obtained with jointly fitted Weibull. Given the latter achieved a slightly better fit with fewer parameters, it was favoured as the optimal fit.

Long-term projections of all-cause and CV mortality using alternative parametric distributions are shown in **Error! Reference source not found.** and Figure 24. **Alternative CV mortality survival curves (scenario analyses)**

respectively. Risk equations derived with alternative distributions for sensitivity analyses are summarised in Table 39. The proportional hazards assumption is assumed to hold and hence regression models jointly fit both arms using treatment as a predictor. Diagnostic plots that evaluated validity of the proportional hazards assumption for all the tested models are shown in Appendix M.

Figure 23. Alternative all-cause mortality survival curves (scenario analyses)

Abbreviations: KM, Kaplan-Meier curve; SoC, standard of care.

Figure 24. Alternative CV mortality survival curves (scenario analyses)

Abbreviations: KM, Kaplan-Meier curve; SoC, standard of care

Table 39. Risk equations for alternative parametric distributions of CV mortality and all-cause mortality (KCCQ-based), ITT population of EMPEROR-Reduced (scenario analyses)

Coefficients	Exponential	Gompertz	Log-normal	Log-logistic	Generalised gamma
All-cause death					
P1					
P2 (intercept)					
P3					
Treatment effect					
KCCQ-CSS: 55 to <75 (Quartile 2)					
KCCQ-CSS: 75 to <90 (Quartile 3)					
KCCQ-CSS: 90 to 100 (Quartile 4)					
CV death					

Coefficients	Exponential	Gompertz	Log-normal	Log-logistic	Generalised gamma
P1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
P2 (intercept)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
P3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Treatment effect	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 55 to <75 (Quartile 2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 75 to <90 (Quartile 3)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 90 to 100 (Quartile 4)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: CSS, clinical summary score; KCCQ, Kansas City Cardiomyopathy Questionnaire.

B.3.3.3.2 Death from non-CV causes

Death attributable to non-CV causes was calculated from the all-cause death and CV death risk equations derived from the EMPEROR-Reduced trial (Section B.3.3.3.1). During each model cycle, the difference between the all-cause death rate and CV death rate was used to calculate non-CV death. If the probability of non-CV death was higher in a given cycle compared to the most recent age- and sex-specific life table probability for the general UK population, the latter was used to inform non-CV death in a given cycle. UK life tables were adjusted to exclude CV-related deaths to avoid double counting. UK life tables used to derive non-CV death for the model are reported in Appendix L.

A scenario analysis was carried out which used the CV death and all-cause death survival curves from EMPEROR-Reduced trial only, without applying the non-CV death rate from UK life tables (Section 0).

B.3.3.4 Incidence of HHF

The monthly rate of first and recurrent HHF was modelled using a Poisson model fitted to patient-level data with generalised estimating equations (GEEs) which had an auto-regressive covariance structure to account for correlations between repeated measures as the data included a record for every month of follow-up for each patient. A negative binomial distribution was also considered but the fitting procedure failed

and produced errors. The fit of the Poisson and negative binomial models were compared without the GEE correction and showed similar fit based on deviance statistics, thus the negative binomial was not pursued further. The HHF rates appeared to be relatively constant over time and the analyses, therefore, assumed a constant rate in each treatment arm, but alternative scenarios where the benefit of treatment is turned off were considered.

The Poisson GEE model included treatment and time-varying KCCQ-CSS health states as predictors and was derived using the ITT population of the EMPEROR-Reduced. The parameters of the fitted GEE used to predict HHF in the ITT population in the model are reported in Table 40.

Table 40. Risk equation for hospitalisation for HF based on Poisson regression, ITT population from EMPEROR-Reduced trial

Covariate	Coefficient	SE	p-value
Intercept	[redacted]	[redacted]	[redacted]
Treatment effect Empagliflozin 10 mg (Ref.: Placebo)	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 55 to <75 (Quartile 2)*	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 75 to <90 (Quartile 3)*	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 90 to 100 (Quartile 4)*	[redacted]	[redacted]	[redacted]

Abbreviations: CSS, clinical summary score; KCCQ, Kansas City Cardiomyopathy Questionnaire; SE, standard error.
* vs KCCQ-CSS: 0 to <55 (Quartile 1)

B.3.3.5 Composite renal outcome

The model has the functionality to capture the benefits of empagliflozin on slowing the decline in renal function associated with the progression of HF. In each cycle, the Markov cohort experiences a risk of a composite renal outcome equivalent to that observed in the EMPEROR-Reduced trial. In line with the trial definition, the composite renal endpoint comprised chronic dialysis, renal transplantation or a sustained eGFR reduction of ≥40% from baseline (for more detail see section B.2.6.2.3) occurring at the rate shown in **Error! Reference source not found.** Once the event occurred, the patient remained in the state “alive with composite renal outcome” for the duration of the model, exiting only upon death, and was assigned a cost and disutility calculated

as a weighted average of costs and disutilities corresponding to renal dialysis and CKD stage 3b [when applied to the mean eGFR at baseline of the EMPEROR-Reduced trial (62.0 ml/min/1.73 m^2), a 40% decline in eGFR corresponds to the CKD stage 3b with a mean eGFR of 37.2 ml/min/1.73 m^2] (Table 41). Since no renal transplants took place during the trial follow-up, the proportion of renal transplants within the composite renal endpoint was zero.

The mortality of patients who had a composite renal outcome was assumed equal to that of the rest of the cohort, i.e. the sum of CV and non-CV death. Also, the proportion of the cohort residing in this state did not influence the distribution of patients across other health states in the model, i.e. the KCCQ states, HHF events, CV and non-CV death. This is a simplistic approach to modelling renal function decline in HF that could be incorporated within the existing Markov structure. Accurate modelling of chronic kidney disease progression would require more complex approaches (184). The simpler approach, however, does not account for the impact of renal decline on cardiovascular outcomes and is therefore conservative with respect to empagliflozin.

Table 41: Clinical inputs composite renal endpoint

	Empagliflozin + SoC	Placebo + SoC	Reference
Incidence rate per 100 years at risk	[redacted]	[redacted]	EMPEROR-Reduced CSR Table 11.1.2.6: 1 (135)
HR vs. placebo	[redacted]	[redacted]	
of which:	[redacted]		
Renal dialysis	[redacted]		
Renal transplant	[redacted]		
Sustained eGFR reduction	[redacted]		

Abbreviations: eGFR, estimated glomerular filtration rate; HR, hazard ratio; SoC, standard of care.

B.3.3.6 Treatment discontinuation

A parametric survival analysis was applied to estimate the time to empagliflozin treatment discontinuation as observed in the EMPEROR-Reduced trial. Analyses considered treatment and time-varying KCCQ-CSS as predictors for discontinuation. Discontinuation due to death was not considered an event in these analyses as these were captured in mortality equations. Patients who died were censored for treatment discontinuation.

The Weibull, log-logistic, log-normal, Gompertz, exponential, and generalised gamma distributions were explored, and fits compared. Diagnostic plots and fit statistic are shown in Appendix M. Fit statistics suggested comparable fit for Weibull, log-normal, log-logistic and generalised gamma (albeit latter with a need for an additional parameter). However, the log-normal, log-logistic and Weibull fit predicted unrealistically long mean time-to-treatment discontinuation of over 14 years. While the exponential distribution appeared to produce a poor fit based on fit statistics and alignment with the observations in the early stages of follow-up, it fits the later portion of the observed curves more closely. It also yielded the more plausible projected mean time-to-treatment discontinuation of approximately seven years. Therefore, exponential distribution was selected as the base-case (Table 42). After discontinuation of empagliflozin + SoC, patients were assumed to receive SoC, and thus experience the same risk of clinical events, costs, and utility decrements as patients on SoC. The treatment discontinuation equations for alternative parametric distributions are provided in Table 43.

Table 42. Risk equation for treatment discontinuation from EMPEROR-Reduced trial, exponential distribution

Covariate	Coefficient	SE	p-value
Log (Scale)	[redacted]	[redacted]	[redacted]
Treatment effect Empagliflozin 10 mg (Ref.: Placebo)	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 55 to <75 (Quartile 2)*	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 75 to <90 (Quartile 3)*	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 90 to 100 (Quartile 4)*	[redacted]	[redacted]	[redacted]

Abbreviations: CSS, clinical summary score; KCCQ, Kansas City Cardiomyopathy Questionnaire; SE, standard error.
* vs KCCQ-CSS: 0 to <55 (Quartile 1)

Table 43. Risk equations for alternative parametric distributions of treatment discontinuation, ITT population from EMPEROR-Reduced

Coefficients	Weibull	Gompertz	Log-normal	Log-logistic	Generalised gamma
P1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
P2 (intercept)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
P3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Treatment effect	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 55 to <75 (Quartile 2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 75 to <90 (Quartile 3)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 90 to 100 (Quartile 4)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: CSS, clinical summary score; KCCQ, Kansas City Cardiomyopathy Questionnaire.

B.3.3.7 Adverse event rates

The risk of experiencing AEs from treatments was informed by the most common AEs of special interest in the EMPEROR-Reduced trial by assuming a constant hazard. The rates of AEs associated with empagliflozin + SoC and SoC were derived from the EMPEROR-Reduced trial. Patients who discontinued empagliflozin were subject to the risk of AEs associated with the placebo arm of the EMPEROR-Reduced trial. The rate of AEs with SGLT2i could not be derived due to different AE definitions and observation periods in EMPEROR-Reduced and DAPA-HF, hence the risk of AEs with an SGLT2i was assumed to be the same as that with empagliflozin (Table 44).

Table 44. Rates of AEs in the modelled cohort

	Rate per 1000 patient years in the EMPEROR-Reduced trial	
	Empagliflozin + SoC	Placebo + SoC
Urinary tract infection	[redacted]	[redacted]
Genital mycotic infection	[redacted]	[redacted]
Acute renal failure	[redacted]	[redacted]
Hepatic injury	[redacted]	[redacted]
Volume depletion	[redacted]	[redacted]
Hypotension	[redacted]	[redacted]
Hypoglycaemic event*	[redacted]	[redacted]
Bone fracture	[redacted]	[redacted]

Abbreviation: SoC, standard of care.
* Defined as an event with a plasma glucose value of ≤70 mg/dL or where assistance was required.

B.3.4 Measurement and valuation of health effects

B.3.4.1 Health-related quality of life data from clinical trials

The purpose of this section is to describe how HRQoL data was collected in the trial and comment on its consistency with Section 5.3 of the reference case (144). The utility and disutility values associated with the model health states, AEs and HHF were obtained from the pooled analysis of the patient-level data for the ITT population in the EMPEROR-Reduced trial (135). The values were derived from responses to the EQ-5D-5L questionnaire collected at baseline and at weeks 12, 32, 52, 100 and 148 following randomisation, at treatment discontinuation, and at a follow-up visit of 30 days following regular or premature completion of the treatment period. Patients' responses to EQ-5D-5L questionnaire were then mapped to EQ-5D-3L descriptive system using the crosswalk mapping function developed by van Hout et al (185). The EQ-5D-3L responses were then converted to utility scores using the published UK utility values for EQ-5D health states, derived with the time trade-off method described by Dolan et. al. (186). Utility scores were analysed with a linear mixed-effects regression to account for the repeated measures on the same patients (187). The model included a random intercept for each patient to account for the within-patient correlations. To capture the short-and long-term effects of HHF events on utilities, the linear mixed model incorporated time-varying indicators reflecting whether a patient had a HHF in 0-1 month, 1-2 months, 2-4 months, and 4-12 months prior versus not hospitalised, as well as time-varying KCCQ-CSS quartiles. The reference group was no HHF events to date, and patients were classified back into the reference group once a year had passed from hospitalisation. This approach allowed estimation of the utilities based on patients' current severity level and HHF status. The AE effects were captured in the same way and assumed to be more acute. Indicators were created for each type of AE to flag whether it had occurred in the previous month and patients were returned to the reference group one month after the AE. The indicators for HHF and AE events were added to the baseline model, which was then trimmed down to

remove predictors that became non-significant. Additionally, the model was adjusted for gender, age, region, and ischaemic cause.

The validity of the linear mixed model approach was verified by assessing the distribution of predicted values from the equations to ensure no ceiling effects were present. Specifically, the predicted values were within the expected ranges and less than 1% of the predicted values were above 1.00 (maximum at 1.004). The utility model used to inform health state utilities and utility decrements associated with clinical events is presented in Table 45.

Table 45. Health-related quality of life equation derived from EMPEROR-Reduced trial

Covariate	Coefficient	SE	t-value
Distribution/Type	**Linear Mixed Model**		
Intercept	[redacted]	[redacted]	[redacted]
Demographics			
Sex: Male	[redacted]	[redacted]	[redacted]
Age ≥65 years	[redacted]	[redacted]	[redacted]
Region			
Region: Asia	[redacted]	[redacted]	[redacted]
Region: Latin America	[redacted]	[redacted]	[redacted]
Region: North America	[redacted]	[redacted]	[redacted]
Region: Other	[redacted]	[redacted]	[redacted]
KCCQ-CSS			
KCCQ-CSS: 55 to <75 (Quartile 2)	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 75 to <90 (Quartile 3)	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 90 to 100 (Quartile 4)	[redacted]	[redacted]	[redacted]
Baseline EQ-5D (standardised)	[redacted]	[redacted]	[redacted]
Medical History			
HF: Ischaemic cause	[redacted]	[redacted]	[redacted]
Time Since HHF	[redacted]	[redacted]	[redacted]

Covariate	Coefficient	SE	t-value
HHF: <1 month	[redacted]	[redacted]	[redacted]
HHF: 1 to <2 months	[redacted]	[redacted]	[redacted]
HHF: 2 to <4 months	[redacted]	[redacted]	[redacted]
HHF: 4 to <12 months	[redacted]	[redacted]	[redacted]
AEs			
Urinary tract infection	[redacted]	[redacted]	[redacted]
Genital mycotic infection	[redacted]	[redacted]	[redacted]
Acute renal failure	[redacted]	[redacted]	[redacted]
Hepatic injury	[redacted]	[redacted]	[redacted]
Volume depletion	[redacted]	[redacted]	[redacted]
Hypotension	[redacted]	[redacted]	[redacted]
Hypoglycaemic event	[redacted]	[redacted]	[redacted]
Bone fracture	[redacted]	[redacted]	[redacted]

Abbreviations: AE, adverse event; CSS, clinical summary score; EQ-5D, EuroQol five dimensions; HF, heart failure; HHF, hospitalisation for heart failure; KCCQ, Kansas City Cardiomyopathy Questionnaire; SE, standard error.

B.3.4.2 Mapping

The patient responses from the EQ-5D-5L questionnaire were mapped to the EQ-5D-3L questionnaire using the crosswalk methodology developed by van Hout et al. with UK value sets (185). This methodology is in line with NICE technology assessment guidelines as indicated in the NICE position statement (188).

B.3.4.3 Health-related quality of life studies

A SLR was conducted to identify relevant utilities evidence related to the treatment of HFrEF. Full details of the process and methods used are described in Appendix H.

A total of 61 studies were included in the HRQoL SLR and all studies are summarised in Appendix H. All studies reported that the EQ-5D was used to measure HRQoL, consistent with the NICE reference case. However, the valuation method was reported in only 14 studies and it varied: eight studies used the time trade-off (TTO), one study used the TTO and visual analogue scale (VAS), four studies used the direct valuation method by the general public and one study used a patient questionnaire to obtain

utility values. Four studies mapped the Minnesota living with heart failure questionnaire (MLwHF) to the EQ-5D. Although the utility values identified in the SLR met the reference case requirements, the economic analysis presented in this submission uses the utility values derived from the EQ-5D data collected in the EMPEROR-Reduced trial from the patient population of direct relevance to the decision problem.

B.3.4.4 Adverse reactions

Utility decrements associated with adverse events of special interest (AESI) were derived either from the trial data as described in B.3.4.1 or from the literature, and applied in the model over the month of incidence only (Table 47). The disutility values for genital mycotic infection, acute renal failure, hepatic injury, volume depletion and bone fracture were derived from patient-level analysis of the EMPEROR-Reduced trial as the trial EQ-5D scores were deemed more reflective of the population of interest than the available estimates from the literature (135). The disutility associated with urinary tract infection was sourced from Sullivan 2016 which provides a catalogue of disutility values for the UK (189). This study reported EQ-5D scores for diabetes-related chronic conditions, based on a nationally representative 12-item Short Form Health Survey response (n=20,705) from the US (189). These responses were mapped to EQ-5D-3L, and subsequently valued using UK-specific EQ-5D tariffs (189). Its multivariate regression model included all diabetes-related comorbidities as independent variables and two comorbidity indexes, and was controlled for region, age, sex, race, ethnicity, education, insurance coverage, family income, and body-mass index (BMI) category. The disutility value for hypotension was assumed equal to that of essential hypertension and taken from literature (190).

B.3.4.5 Composite renal endpoint disutility

An annual disutility of -[redacted] was applied to the cohort residing in the state “alive with a composite renal endpoint” throughout the time horizon and for as long as patients

remained alive in this state. The value was calculated as a weighted average of disutilities associated with dialysis and CKD stage 3 using a fixed ratio of dialysis and sustained eGFR reduction events in EMPEROR-Reduced (Table 46). The disutilities associated with the clinical events within the composite endpoint were sourced from the study of Jesky et al. (191). The disutility of renal dialysis was calculated as the difference between utilities of CKD stage 5 and stages 1-2, while that of sustained eGFR reduction was derived as the differences between CKD stages 3 and 1-2, thus accounting for the marginal loss in quality of life associated with the respective stages of renal decline.

Table 46. Disutility associated with the composite renal endpoint

Event within the composite endpoint	Mean	Reference
Dialysis	-0.12 = 0.73 - 0.85	Jesky et al (191)
Renal transplant	-0.12 = 0.73 - 0.85	
Sustained eGFR reduction	-0.05 = 0.80 – 0.85	
Weighted average composite renal	[redacted]	Calculated using fixed weights shown in Table 41

Abbreviations: eGFR, estimated glomerular filtration rate.

B.3.4.6 HRQoL experienced in each health state

KCCQ-CSS health state-specific utility values were derived from the HRQoL equation shown in Table 44 based on EQ-5D data from EMPEROR-Reduced (135). The derived mean utility values were adjusted for gender, age (≥65 years), geographical region (Asia, Latin America, North America), baseline EQ-5D, and medical history (ishaemic cause of HF and history of HHF). The estimation of QALYs was determined by KCCQ-CSS state occupancy over time, and the incidence of discrete clinical events such as HHF and AEs (urinary tract infection, genital mycotic infection, acute renal failure, hepatic injury, volume depletion, hypotension, hypoglycemic event, and bone fracture) which were captured as one-off utility decrements for the proportion of cohort experiencing the event in the month of incidence. No change in health state utility was considered based on age. This was a simplifying assumption considering the short life expectancy of patients with HFrEF.

B.3.4.7 HRQoL over the course of the disease

Changes in HRQoL over the course of the model were fully accounted for by changes in KCCQ-CSS health state occupancy, HHF events and occurrence of AEs, which were included in the model as time-varying predictors (Sections B.3.4.1 and B.3.4.5). There were no statistically significant differences in EQ-5D scores between the two treatment groups in the EMPEROR-Reduced trial, hence the treatment was not a predictor in the utility equation.

B.3.4.8 Baseline HRQoL

The baseline utility values were contingent on the KCCQ-CSS quartile health states and were derived from the EMPEROR-Reduced data.

B.3.4.9 Adjusted health state utility values

As the trial-derived utility value for KCCQ-CSS quartile 4 [redacted]) was higher than the utility of UK general population aged 60 to 69 years (0.7740) reported by Sullivan et al. 2011, an age-adjustment was applied (192). Under this adjustment, utility values for KCCQ-CSS quartile 1–3 were reduced by the relative difference between EMPEROR-Reduced observed utility for KCCQ-CSS quartile 4 and published utility of UK general population aged 60 to 69 years (192). The latter was assumed as the utility value for KCCQ-CSS quartile 4 utility in the model. The model provides flexibility to exclude the age-adjustment factor and instead use the EMPEROR-Reduced derived health state utility values, which was considered in a scenario analysis (135).

B.3.4.10 Summary of utility values

The health state utility values and disutilities associated with clinical events and AEs included in the cost-effectiveness analysis are outlined in Table 47. Health state utility values and HHF disutility values were derived from the linear mixed-effects regression of EQ-5D data from the EMPEROR-Reduced trial. The AE disutility values were either derived from the trial data or identified from targeted literature searches.

Table 47. Summary of utility values for cost-effectiveness analysis

Parameter	Mean Utility	SE	Source and justification	Reference in the submission
Without Age-adjustment				
KCCQ-CSS: 0 to <55 (Quartile 1)	[redacted]	[redacted]	Based on EMPEROR-Reduced trial data analyses (135)	B.3.4.1 and Table 45
KCCQ-CSS: 55 to <75 (Quartile 2)	[redacted]	[redacted]		B.3.4.1and Table 45
KCCQ-CSS: 75 to <90 (Quartile 3)	[redacted]	[redacted]		B.3.4.1and Table 45
KCCQ-CSS: 90 to 100 (Quartile 4)	[redacted]	[redacted]		B.3.4.1and Table 45
With Age-adjustment (base-case)				[redacted]
KCCQ-CSS: 0 to <55 (Quartile 1)*	[redacted]	[redacted]	Based on EMPEROR-Reduced trial data analyses (135)	B.3.4.1, Table 45 and B.3.4.9
KCCQ-CSS: 55 to <75 (Quartile 2)*	[redacted]	[redacted]		B.3.4.1, *Table 45* and B.3.4.9
KCCQ-CSS: 75 to <90 (Quartile 3)*	[redacted]	[redacted]		B.3.4.1, Table 45 and B.3.4.9
KCCQ-CSS: 90 to 100 (Quartile 4)^	[redacted]	[redacted]		B.3.4.1, Table 45 and B.3.4.9
Clinical Event Disutility				[redacted]
HHF	[redacted]	[redacted]	Based on EMPEROR-Reduced trial data analyses (135)	B.3.4.1 and Table 45
AE Disutilities				
Urinary tract infection	[redacted]	[redacted]	Sullivan 2016 (189)	B.3.4.1, Table 45 and B.3.4.4
Genital mycotic infection	[redacted]	[redacted]	Based on EMPEROR-Reduced trial data analyses (135)	B.3.4.1, Table 45 and B.3.4.4
Acute renal failure	[redacted]	[redacted]		B.3.4.1, Table 45 and B.3.4.4

Parameter	Mean Utility	SE	Source and justification	Reference in the submission
Hepatic injury	[redacted]	[redacted]		B.3.4.1, Table 45 and B.3.4.4
Volume depletion	[redacted]	[redacted]		B.3.4.1, Table 45 and B.3.4.4
Hypoglycaemic event	[redacted]	[redacted]		B.3.4.1, Table 45 and B.3.4.4
Hypotension^^	–0.025	[redacted]	Sullivan 2006 (190)	B.3.4.4
Bone fracture	[redacted]	[redacted]	Based on EMPEROR-Reduced trial data analyses (135)	B.3.4.1, Table 45 and B.3.4.4

Abbreviations: AE, adverse event; CV, cardiovascular; CSS, clinical summary score; HHF, hospitalisation due to heart failure; KCCQ, Kansas City Cardiomyopathy Questionnaire; SE, standard error

* Relative differences from EMPEROR-Reduced study applied to general population utility for people aged 60 to 69 years reported by Sullivan et al. 2011 (192)

^ Set equal to UK general population utility for people aged 60 to 69 years reported by Sullivan et al. 2011 (192)

^^ Disutility for hypertension in the US population reported by Sullivan 2006 due to lack of UK population values (190)

B.3.5 Cost and healthcare resource use identification, measurement, and valuation

A SLR was conducted to identify cost and healthcare resource use evidence related to the treatment of HFrEF. Full details of the process and methods are provided in Appendix I. In summary, a total of 14 studies, 12 from the UK and two international studies were identified. Their brief overview is provided in Appendix I.

B.3.5.1 Intervention and comparators' costs and resource use

Intervention and comparator costs in the model consist of drug acquisition costs and monitoring costs. In base-case analysis, empagliflozin + SoC is compared to SoC alone. A scenario analysis considers SGLT2i class of drugs as addition to SoC relative to SoC alone. The drug costs for empagliflozin and the SoC therapies were extracted from the NHS Electronic Drug Tariff (193).

Patients were assumed to receive appropriately titrated doses of SoC therapies, (i.e. the stable maintenance dosage for each SoC treatment was applied and the titration process was not modelled). The background therapy used within the trial was as per national or international guideline recommendations (20, 34). Costs of devices were not included as patients were assumed to have undergone procedures for these treatments before entering the model.

A summary of the pack cost, pack size, strength, dosage, daily and monthly cost are provided in Table *48*.

Table 48: Technology and comparator unit costs

Drug class	Treatment	Pack cost (MIMS)	Pack size	Strength (mg)	Daily dosage	Daily cost	Monthly cost	Source
SGLT2i	**Empagliflozin**	£36.59	28 pills	10 mg	10 mg	£1.31	£39.78	NHS Electronic Drug Tariff (193).
	Dapagliflozin	£36.59	28 pills	10mg	10mg	£1.31	£39.78	
ARNi	**Sacubitril valsartan**	£91.56	56 pills	200 mg	400 mg	£3.92	£119.44	
Loop diuretics	**Furosemide**	£0.94	28 pills	40mg	80 mg	£0.07	£2.04	
ACEi	**Captopril**	£1.68	56 pills	50 mg	100 mg	£0.06	£1.83	
	Enalapril	£13.35	28 pills	20 mg	20 mg	£0.48	£14.51	
	Lisinopril	£1.19	28 pills	20 mg	20 mg	£0.04	£1.29	
	Ramipril	£1.42	28 pills	10 mg	10 mg	£0.05	£1.54	
	Trandolapril	£1.68	14 pills	0.5 mg	1.5 mg	£0.36	£10.96	
BB	**Bisoprolol**	£1.07	28 pills	10 mg	10 mg	£0.04	£1.16	
	Carvedilol	£1.99	28 pills	25 mg	50 mg	£0.14	£4.43	
	Metoprolol	£2.51	28 pills	100 mg	100 mg	£0.09	£2.73	
	Nebivolol	£27.39	28 pills	10 mg	10 mg	£0.98	£29.77	
ARBs	**Candesartan**	£1.93	28 pills	32 mg	32 mg	£0.07	£2.10	
	Valsartan	£12.98	28 pills	160 mg	320 mg	£0.93	£28.22	
	Losartan	£1.73	28 pills	100 mg	150 mg	£0.09	£2.82	
MRA	**Eplerenone**	£7.02	28 pills	50 mg	50 mg	£0.25	£7.63	

Abbreviations: ACEi, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor-blockers; ARNi, angiotensin receptor-neprilysin inhibitor; BB, beta-blockers; MRA, mineralocorticoid receptor antagonist; SGLT2i, sodium-glucose co-transporter-2 inhibitor

B.3.5.2 Health state unit costs and resource use

The health state resource use and unit costs associated with HFrEF are outlined in Table 49 and Table 50. The costs of managing clinical events were obtained from UK national databases or published literature.

The acute cost of HHF was based on NHS reference costs for non-elective long inpatient stay, computed as the weighted average of reference costs for healthcare resource group (HRG) codes EB03A to EB03E and the number of finished consultant episodes (FCEs) (194).

The cost of CV death was estimated from a study by Alva et al., who estimated inpatient costs for T2D-related complications during the UK Prospective Diabetes Study post-trial monitoring period from 1997 to 2007 using hospitalisation records for patients in England (n=2,791) (195). Their analysis produced an equation with coefficients interpreted as linear effects on expected inpatient costs for complications, which was used in the cost-effectiveness analysis to estimate costs of a fatal myocardial infarction, fatal ischaemic heart disease, and fatal stroke for a male aged <65 years and ≥65 years and a female aged <65 years and ≥65 years. Characteristics of EMPEROR-Reduced participants (e.g., percentage of male or female and percentage of aged <65 or ≥65 years) were applied to derive weighted average costs for each event, which were themselves averaged to derive the cost of CV death for the model. Non-CV deaths were assumed to incur no cost in the base-case and were assumed to equal the cost of CV death in a scenario analysis (135).

A weighted average annual cost associated with the composite renal outcome was based on published costs of the individual renal outcomes (i.e., dialysis, renal transplantation, and sustained eGFR reduction) and the percentage of patients in the EMPEROR-Reduced trial affected by each (12.5%, 0%, and 87.5%, respectively) among those experiencing the composite. A study by Kerr et al. provided an estimate of the mean annual cost to the English NHS (2009-2010 prices) of direct chronic kidney disease (CKD) care per patient on dialysis and per transplant recipient (196). To estimate the cost of sustained eGFR reduction, a 40% eGFR decline was applied to

the mean eGFR at baseline in the EMPEROR-Reduced trial (62.0 ml/min/1.73 m^2), resulting in an eGFR value of 37.2 ml/min/1.73 m^2 (i.e., CKD stage 3b). The unit cost for CKD stage 1-3B (£511.23, inflated to 2021 prices) was obtained from the study by Kent et al who estimated annual UK hospital care costs by CKD stage (2010-2011 prices) based on analyses of the Study of Heart and Renal Protection (SHARP) trial which prospectively collected information on kidney disease progression in a cohort of patients with moderate to severe CKD (197). All CV and renal event costs were inflated to 2021 by applying the consumer price health inflation factor from Eurostat (1.01) where applicable (see Appendix N) (198). The cost of composite renal outcome assumes a fixed ratio of frequency of dialysis relative to eGFR decline, such that it does not reflect the possibility of further progression of eGFR to end-stage renal disease. This is a conservative scenario that does not favour empagliflozin beyond the renal effects observed in EMPEROR-Reduced.

The HF-related disease management costs associated with GP, cardiologist visits and A&E referral were computed based on the frequency of use and unit cost for each type of care. Resource use was based on data from the Clinical Practice Research Datalink, as reported by McMurray and colleagues (2018), which was converted from annual to monthly frequency (166). Unit costs were retrieved from national sources. In particular, the cost of GP and cardiologist visits were based upon per patient contact lasting 9.22 minutes (code 10.3b) and a consultant-led non-admitted face to face follow-up appointment in cardiology (code 320), respectively, while A&E referral cost was a weighted mean derived from national average unit costs and number of FCEs for non-admitted emergency medicine (codes VB01Z to VB11Z, and VB99Z) (194, 199). All disease management costs were inflated to 2021 by applying the consumer price health inflation factor from Eurostat (198).

Table 49. CV and renal events frequencies and management unit costs

CV and renal events	Unit cost per event (inflated to 2021)	Event rate (per 1,000 patient years)**				Cost source
		HFrEF with T2DM		HFrEF without T2DM		
		SoC	Empagliflozin + SoC	SoC	Empagliflozin + SoC	
HHF	£3,071.65	[redacted]	[redacted]	[redacted]	[redacted]	NHS 2018-2019 (194); Weighted average of non-elective long stay HRG codes; EB03A: EB03E
CV death	£4,146.38	[redacted]	[redacted]	[redacted]	[redacted]	Alva 2015 (195)
Non-CV death*	£0.00	[redacted]	[redacted]	[redacted]	[redacted]	Assumption
Composite renal outcome***	£4,862.38	[redacted]	[redacted]	[redacted]	[redacted]	Weighted costs of the following: Dialysis (196); Renal transplantation (196); Sustained eGFR reduction (197)

* Non-CV death is used in model background calculations to correctly compute the number of patients remaining alive (and on-treatment) from year to year.
** Source: EMPEROR-REDUCED Clinical trials report (135).
*** Composite renal outcome is defined as chronic dialysis, renal transplantation, or a sustained reduction of ≥ 40% in the eGFR or a sustained eGFR <15ml/min/1.73m^2 in patients with a baseline eGFR of ≥30 ml /min/1.73 m^2 or a sustained eGFR <10 ml/min/1.73 m^2 in those with a baseline eGFR of <30 ml/min/1.73 m^2.
Abbreviations: eGFR, estimated glomerular filtration rate; CV, cardiovascular; HHF, hospitalisation for heart failure; HRG, healthcare resource group

Table 50. Disease management resource use and unit costs

Disease management costs	Unit cost per event	Monthly frequency for all KCCQ-CSS quartiles*	Monthly cost per patient for all KCCQ-CSS quartiles	Cost source
GP visit	£39.73	1.928	£76.62	PSSRU 2020, Code 10.3b (9.22 minutes per patient contact) (199)
Cardiologist visit	£140.00	0.004	£0.59	NHS 2018-2019, Cardiology non-admitted face to face and follow-up visit (194)
A&E referral	£153.60	0.008	£1.23	NHS 2018-2019, weighted mean (HRG codes: VB01Z-VB11Z, VB99Z (194))
Total cost			**£78.43**	

Abbreviations: A&E, accident and emergency; HRG, healthcare resource group; PSSRU, personal social services research unit
*Monthly frequency: McMurray et al (2018) (166)

B.3.5.3 Adverse reaction unit costs and resource use

The adverse reaction unit costs and resource use are provided Table 51 below. The acute cost of an outpatient visit was based on costs for general practitioners (GP), assuming per patient contact lasting 9.22 minutes (code 10.3b), taken from unit costs of health and social care by personal social services research unit (PSSRU) 2020 (199). In addition, NHS reference costs for non-elective long and short stays for HRG codes related with each AE served as the basis for the cost of inpatient episodes (Table 51), while self-treated patients were assumed to receive over-the-counter treatment, thus incurring no costs to the health care payer (194). The distribution of visit types for management of AEs was based on assumption, as UK-specific data was not available and the event rate was derived from the EMPEROR-Reduced trial (135). All AE costs were inflated to 2021 by applying the consumer price health inflation factor from Eurostat (198).

Table 51: Adverse event management unit costs, event rate and frequencies of distribution

Adverse event management costs	Weighted average cost (2021)	Outpatient*	Inpatient	Event rate for overall HFrEF (per 1,000 patient years)**		Inpatient cost source
		Unit cost (inflated to 2021) (%)	Unit cost (inflated to 2021) (%)	SoC	Empagliflozin + SoC	
UTI	£39.73	£39.73 (100%)	£1,670.95 (0%)	[redacted]	[redacted]	NHS 2018-2019 (194); weighted average of HRG codes: LA04H, LA04J-N, LA04P-S, kidney or urinary tract infections, non-elective long or short stay
GMI	£39.73	£39.73 (100%)	£1,133.48 (0%)	[redacted]	[redacted]	NHS 2018-2019 (194); weighted average of HRG codes: WJ03A-G, standard infection diseases, non-elective long or short stay
Acute renal failure	£1,905.51	£39.73 (0%)	£1,905.51 (100%)	[redacted]	[redacted]	NHS 2018-2019 (194); weighted average of HRG codes: LA07H, LA07J-N, LA07P, acute kidney injury, non-elective long or short stay
Hepatic injury	£1,273.77	£39.73 (50%)	£2,507.80 (50%)	[redacted]	[redacted]	NHS 2018-2019 (194); weighted average of HRG codes: GC01C-F, liver failure disorders, non-elective long or short stay
Volume depletion	£39.73	£39.73 (100%)	£1,361.45 (0%)	[redacted]	[redacted]	NHS 2018-2019 (194); weighted average of HRG codes: KC05G-H, KC05J-N, fluid or electrolyte disorders, non-elective long or short stay
Hypotension	£39.73	£39.73 (100%)	£1,807.11 (0%)	[redacted]	[redacted]	NHS 2018-2019 (194); weighted average of HRG codes: EB14A-E, other acquired cardiac conditions, non-elective long or short stay
Hypoglycaemic event***	£626.54	£39.73 (50%)	£1,213.35 (50%)	[redacted]	[redacted]	NHS 2018-2019 (194); weighted average of HRG codes: KA08A-C, other endocrine disorders, non-elective long or short stay
Bone fracture	£2,709.96	£39.73 (0%)	£2,709.96 (100%)	[redacted]	[redacted]	NHS 2018-2019 (194); weighted average of HRG codes: HD39D-H, pathological fractures, non-elective long or short stay

*Outpatient source: PSSRU 2020, Code 10.3b (9.22 minutes per patient contact) and inflated to 2021 (199)
**Event rate source: EMPEROR-REDUCED Clinical trials report (135)
***Defined as event with a plasma glucose value of ≤70 mg/dL or where assistance was required.
Abbreviations: GMI, Genital mycotic infection; HFrEF, heart failure and reduced ejection fraction; HRG, healthcare resource group; UTI, Urinary tract infection

B.3.5.4 Miscellaneous unit costs and resource use

There are no miscellaneous unit costs and resource use.

B.3.6 Summary of base-case analysis inputs and assumptions

B.3.6.1 Summary of base-case analysis inputs

Table 52. Summary of variables applied in the economic model

Variable	Value	SE	Distribution	Reference
Baseline characteristics				
Age (years)	66.84	0.18	NA	Table 36
Male	76%	0.01	NA	
KCCQ-CSS Q1: 0-<55.2	[redacted]	[redacted]	NA	
KCCQ-CSS Q2: 55.2-<75	[redacted]	[redacted]	NA	
KCCQ-CSS Q3: 75-<89.6	[redacted]	[redacted]	NA	
KCCQ-CSS Q4: 89.6-100	[redacted]	[redacted]	NA	
Ischaemic HF	51.7%	0.01	NA	
Treatment use at baseline			NA	
ACEi	45.4%	0.01	NA	
ARB	24.3%	0.01	NA	
ARNi	19.5%	0.01	NA	
MRA	71.3%	0.01	NA	
BB	94.7%	0.00	NA	
IVA	0%	0.00	NA	
Loop or High ceiling Diuretics	84.5%	0.00	NA	
Cardiac glycosides	0.0%	0.00	NA	
Nitrates	0.0%	0.00	NA	
Hydralazine	0.0%	0.00	NA	
Monthly KCCQ-CSS transition matrix – months 0–3, Empagliflozin + SoC				
KCCQ [1,1]	[redacted]	[redacted]	Dirichlet	Table 36
KCCQ [1,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [1,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [1,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,1]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,1]	[redacted]	[redacted]	Dirichlet	

Variable	Value	SE	Distribution	Reference
KCCQ [3,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,1]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,4]	[redacted]	[redacted]	Dirichlet	
Monthly KCCQ-CSS transition matrix – months 4-8, Empagliflozin + SoC				
KCCQ [1,1]	[redacted]	[redacted]	Dirichlet	Table 36
KCCQ [1,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [1,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [1,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,1]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,1]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,1]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,4]	[redacted]	[redacted]	Dirichlet	
Monthly KCCQ-CSS transition matrix – months 9+, Empagliflozin + SoC				
KCCQ [1,1]	[redacted]	[redacted]	Dirichlet	Table 36
KCCQ [1,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [1,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [1,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,1]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,1]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,1]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,4]	[redacted]	[redacted]	Dirichlet	
Monthly KCCQ-CSS transition matrix – months 0–3, Placebo + SoC				
KCCQ [1,1]	[redacted]	[redacted]	Dirichlet	Table 36
KCCQ [1,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [1,3]	[redacted]	[redacted]	Dirichlet	

Variable	Value	SE	Distribution	Reference
KCCQ [1,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,1]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,1]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,1]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,2]	0.009	0.0009	Dirichlet	
KCCQ [4,3]	0.096	0.0096	Dirichlet	
KCCQ [4,4]	0.889	0.0889	Dirichlet	
Monthly KCCQ-CSS transition matrix – months 4-8, Placebo + SoC				
KCCQ [1,1]	[redacted]	[redacted]	Dirichlet	Table 36
KCCQ [1,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [1,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [1,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,1]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,1]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,1]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,4]	[redacted]	[redacted]	Dirichlet	
Monthly KCCQ-CSS transition matrix – months 9+, Placebo + SoC				
KCCQ [1,1]	[redacted]	[redacted]	Dirichlet	Table 36
KCCQ [1,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [1,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [1,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,1]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [2,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,1]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [3,4]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,1]	[redacted]	[redacted]	Dirichlet	

Variable	Value	SE	Distribution	Reference
KCCQ [4,2]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,3]	[redacted]	[redacted]	Dirichlet	
KCCQ [4,4]	[redacted]	[redacted]	Dirichlet	
Adjusted CV mortality survival equation* (Weibull)				
Shape	[redacted]	[redacted]	Multivariate normal	Table 38
Scale	[redacted]	[redacted]	Multivariate normal	
Treatment effect	[redacted]	[redacted]	Multivariate normal	
KCCQ Q2	[redacted]	[redacted]	Multivariate normal	
KCCQ Q3	[redacted]	[redacted]	Multivariate normal	
KCCQ Q4	[redacted]	[redacted]	Multivariate normal	
Adjusted all-cause mortality survival equation* (Weibull)				
Shape	[redacted]	[redacted]	Multivariate normal	Table 38
Scale	[redacted]	[redacted]	Multivariate normal	
Treatment effect	[redacted]	[redacted]	Multivariate normal	
KCCQ Q2	[redacted]	[redacted]	Multivariate normal	
KCCQ Q3	[redacted]	[redacted]	Multivariate normal	
KCCQ Q4	[redacted]	[redacted]	Multivariate normal	
Adjusted generalised estimating equations for HHF events* (Poisson)				
Intercept	[redacted]	[redacted]	Multivariate normal	Table 40
Treatment effect	[redacted]	[redacted]	Multivariate normal	
KCCQ Q2	[redacted]	[redacted]	Multivariate normal	
KCCQ Q3	[redacted]	[redacted]	Multivariate normal	
KCCQ Q4	[redacted]	[redacted]	Multivariate normal	
Treatment discontinuation equations* (Exponential)				
Log (Scale)	[redacted]	[redacted]	Multivariate normal	**Table *42***
Treatment effect	[redacted]	[redacted]	Multivariate normal	
KCCQ Q2	[redacted]	[redacted]	Multivariate normal	
KCCQ Q3	[redacted]	[redacted]	Multivariate normal	
KCCQ Q4	[redacted]	[redacted]	Multivariate normal	
Adverse events rates 1-month per cycle – Empagliflozin + SoC				
Urinary tract infection	[redacted]	[redacted]	Gamma	Table 44
Genital Mycotic Infection	[redacted]	[redacted]	Gamma	
Acute renal failure	[redacted]	[redacted]	Gamma	
Hepatic injury	[redacted]	[redacted]	Gamma	
Volume depletion	[redacted]	[redacted]	Gamma	
Hypotension	[redacted]	[redacted]	Gamma	
Hypoglycemic event	[redacted]	[redacted]		
Bone fracture	[redacted]	[redacted]	Gamma	
Adverse events rates 1-month per cycle – Placebo + SoC				
Urinary tract infection	[redacted]	[redacted]	Gamma	Table 44
Genital Mycotic Infection	[redacted]	[redacted]	Gamma	

Variable	Value	SE	Distribution	Reference
Acute renal failure	[redacted]	[redacted]	Gamma	
Hepatic injury	[redacted]	[redacted]	Gamma	
Volume depletion	[redacted]	[redacted]	Gamma	
Hypotension	[redacted]	[redacted]	Gamma	
Hypoglycemic event	[redacted]	[redacted]	Gamma	
Bone fracture	[redacted]	[redacted]	Gamma	
Utility values – health states and events				
KCCQ Q1	[redacted]	[redacted]	Beta	Table 45
KCCQ Q2	[redacted]	[redacted]	Beta	
KCCQ Q3	[redacted]	[redacted]	Beta	
KCCQ Q4	[redacted]	[redacted]	Beta	
HHF (decrement)	[redacted]	[redacted]	Beta	
Disutility values – adverse events				
Urinary tract infection	[redacted]	[redacted]	Beta	Table 46
Genital Mycotic Infection	[redacted]	[redacted]	Beta	
Acute renal failure	[redacted]	[redacted]	Beta	
Hepatic injury	[redacted]	[redacted]	Beta	
Volume depletion	[redacted]	[redacted]	Beta	
Hypotension	[redacted]	[redacted]	Beta	
Hypoglycemic event	[redacted]	[redacted]	Beta	
Bone fracture	[redacted]	[redacted]	Beta	
Treatment acquisition costs per cycle				
Empagliflozin + SoC	82.66	16.53	N/A	Table 48
SoC	42.88	8.58	N/A	
Health state and event costs				
HHF	3071.65	614.33	Gamma	Table 50
CV death	4146.38	829.28	Gamma	
Non-CV death	0.000	N/A	N/A	
Composite renal endpoint costs				
Annual state cost	4862.38	972.48	Gamma	Table 49
Adverse event unit costs				
Urinary tract infection	39.732	7.946	Gamma	Table 51
Genital Mycotic Infection	39.732	7.946	Gamma	
Acute renal failure	1905.51	381.10	Gamma	
Hepatic injury	1273.77	254.75	Gamma	
Volume depletion	39.73	7.95	Gamma	
Hypotension	39.73	7.95	Gamma	
Hypoglycemic event	626.54	125.31	Gamma	
Bone fracture	2709.96	541.99	Gamma	
Resource use costs: Monitoring costs per cycle				
GP visit	39.73	7.95	Gamma	

Variable	Value	SE	Distribution	Reference
Cardiologist visit	140.00	28.00	Gamma	Table 50
A&E referral	153.60	30.72	Gamma	
KCCQ Q1	77.33	15.47	Gamma	
KCCQ Q2	77.33	15.47	Gamma	
KCCQ Q3	77.33	15.47	Gamma	
KCCQ Q4	77.33	15.47	Gamma	
Hazard ratios of SGLT2i vs Empagliflozin + SoC used in scenario analysis				
HR of all-cause mortality	0.946	1.1161	Log-normal	Derived from Zannad et al., 2020 (119) using the corresponding HRs of empagliflozin + SoC vs SoC and SLGT2i vs SoC
HR of CV mortality	0.935	1.1288	Log-normal	
HR of HHF	1.000	1.1050	Log-normal	
First kidney composite outcome	1.192	1.4185	Log-normal	

Abbreviations: A&E, Accident & Emergency; ACEi, Angiotensin converting enzyme inhibitors; ARB, Angiotensin receptor blockers; ARNi, angiotensin receptor neprilysin inhibitor; BB, beta-blocker; CI, confidence interval; CSS, clinical summary score; CV, Cardiovascular; GP, General Practitioner; HHF, hospitalisation for Heart Failure; IVA, ivabradine; KCCQ, Kansas City Cardiomyopathy Questionnaire; MRA, Mineralocorticoid receptor antagonists; SE, standard error; SoC, Standard of Care, SGLT2i, Sodium-glucose co-transporter-2 inhibitors; HR, Hazard Ratio.
*The standard errors reported here are different from the standard errors obtained directly from regression models due to Cholesky decomposition.

B.3.6.2 Assumptions

The cost-effectiveness model uses best the available evidence to inform the decision problem, in line with the NICE reference case and guidance on methods of appraisal (8). Table 53 outlines the simplifying assumptions which were needed either in the absence of data or to ensure internal validity (i.e. preserve alignment between the modelled outcomes and those observed in the EMPEROR-Reduced trial).

Table 53. Summary of key assumptions of the economic analysis

#	Assumption	Justification	Likely bias direction
Model structure			
1	Clinical event rates observed in clinical practice mirror those observed in the EMPEROR-Reduced trial	Generalisability of trial outcomes to clinical practice is a common assumption in economic modelling. In this case, the plausibility of the assumption is strengthened by the trial protocol requirement for consistency with local guidelines in standards of HF care. Thus, event rates observed in the trial have direct relevance to clinical practice.	None
2	There may be unmodelled comorbidities that could have	This choice was made to ensure that the model retains internal validity i.e. the model-	None

#	Assumption	Justification	Likely bias direction
	influenced the shapes of the statistical extrapolations for HHF and death (CV or all-cause). The risk equations incorporated only treatment and time-varying KCCQ-CSS states as predictors	predicted outcomes match those observed in the EMPEROR-Reduced trials (see section B.2.4.1).	
3	The rate of clinical events beyond trial duration is based on extrapolation of the observed trial outcomes	This is a limitation inherent to most cost-effectiveness models. No reliable external source was available for estimating the rate of clinical events beyond the duration of the trial. An external validation exercise was undertaken to assess the validity of the long-term extrapolations (see section B.3.10)	None
4	The HHF risk equations/rates only account for non-fatal HHF events	Fatal HHF events were captured by the CV death risk equations.	None
5	Patients can experience non-CV death in any KCCQ-CSS health state. Non-CV death was computed separately	This is a commonly utilised approach in cost-effectiveness analysis to model death from other causes.	None
6	Temporal changes in serum concentration of NT-pro-BNP, a prognostic biomarker of HF morbidity and mortality, were not modelled	This is unlikely to affect cost-effectiveness results since risk equations account for the time-updated KCCQ-CSS health states and therefore capture the impact of disease severity on all-cause mortality, CV mortality, and the risk of HHF.	None
7	The model assumed that all patients receive appropriately titrated doses of HF medications (e.g., ACEi/ARB). This does not imply that patients receive the maximum effective dose but a dosage that achieves the best trade-off between effectiveness and tolerability	This is a simplifying assumption not expected to lead to bias in favour of any treatment.	None
8	The modelled rate of treatment discontinuation is derived from the EMPEROR- Reduced trial, with a rate of discontinuation applied to all patients receiving empagliflozin + SoC in each modelled cycle (based on the selected distribution). Following discontinuation of empagliflozin, patients are assumed to have the same event risks and costs as patients in the control arm	Patients were on SoC before receiving add-on empagliflozin and it is reasonable to assume that they will continue SoC after discontinuing empagliflozin.	None
9	Empagliflozin can delay the progression of renal disease compared to SoC	To appropriately capture the capacity of empagliflozin to slow the progression of renal impairment associated with chronic HF and cardiorenal syndrome, a composite renal outcome was included in the model (see B.3.3.5.). This includes disutility and cost; however, the impact of change in renal composite outcome over time on survival was	

#	Assumption	Justification	Likely bias direction
		not fully captured as it would introduce additional complexity in the CE model.	
HRQoL			
10	KCCQ-CSS health state-specific utility values and disutilities associated with AEs and HHF were derived from pooled analysis of the EMPEROR-Reduced ITT population after mapping EQ-5D-5L data to EQ-5D-3L and applying the UK value sets	This is common practice when EQ-5D data is available from the clinical trial and in line with the NICE reference case (144)	None
11	The model assumed no decline in HRQoL with increasing age	This is a simplifying assumption considering the short life expectancy of patients with HFrEF. The model includes functionality to adjust utilities to reflect those of age-matched UK general population and in that way already partially reflects the expected utility for the corresponding age groups (see sections B.3.4.8 and B.3.4.9)	None
Costs and resource use			
12	The cost of non-CV death was assumed to be zero under all interventions	The rate of non-CV death is expected to be the same across all interventions, hence this assumption is unlikely to have an impact on incremental cost-effectiveness. Nevertheless, a non-zero cost was provided to allow testing of the alternative costing scenario.	None
13	The model does not include the cost of medical devices or their implantation. It is assumed that patients with ICD/CRT had the device implanted prior to entering the model	This is a simplifying assumption and is not expected to bias any treatment.	None
14	Costs of recognised but relatively mild AEs (e.g., polyuria, episodes of dehydration) associated with SGLT2i are not incorporated in the model	These events generally lead to short episodes of care and low management costs and are expected to have a minimal impact on overall healthcare costs that would not materially affect the model findings.	None
15	The model does not capture episodes of diabetic ketoacidosis, a rare complication associated with SGLT2 inhibition	No cases of ketoacidosis were observed in the EMPEROR-Reduced trial; this assumption is further supported by the EMPA-REG OUTCOME trial in patients with T2DM and established CV disease where very few cases were observed and there was no imbalance between treatment groups (163)	None

Abbreviations: AEs, adverse events; ACEi, Angiotensin converting enzyme inhibitors; ARB, Angiotensin receptor blockers; NT-pro-BNP, N-terminal pro hormone B-type natriuretic peptide; CV, cardiovascular; CRT, cardiac resynchronisation therapy; EQ-5D-5L, EuroQol-5 dimensions-5 levels; EQ-5D-3L, EuroQol-5 dimensions-3 levels; HHF, hospitalisation for heart failure; HFrEF, heart failure with reduced ejection fraction; HRQoL, health-related quality of life; ICD, implantable cardioverter-defibrillator; ITT, intention to treat; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire Clinical Summary Score; NICE, National Institute for

Health and Care Excellence; NT-pro-BNP, N-terminal pro T2DM, type 2 diabetes mellitus; SGLT2i, sodium-glucose co-transporter 2 inhibitor; UK, United Kingdom.

B.3.7 Base-case results

B.3.7.1 Base-case incremental cost-effectiveness analysis results

Table 54 shows the discounted results of the base-case comparison of empagliflozin as an add-on to standard care (ACEi/ARB + BB ± MRA) against SoC alone over a lifetime horizon. SoC is associated with 5.83 life years, 3.78 QALYs, and £17,950 per patient. Treatment with empagliflozin as an add-on to SoC resulted in an increase in life years (+0.21 per person) and QALYs (+0.22) per person at an additional cost of £1,063 per person. Empagliflozin as an add-on to SoC was cost-effective against SoC alone at usual threshold values with an ICER of £4,804 (ICER was <£10k) per QALY gained. The similarity of the overall incremental LY and QALY estimates reflects the fact that patients do not always gain LYs across all KCCQ quartiles, and that the LY differences in each KCCQ state are associated with different utility values.

The main driver of incremental costs associated with the empagliflozin + SoC treatment was the additional cost of empagliflozin, which was partially offset by cost-savings from reduced incidence of HHF and CV death. The incremental QALY gains were driven by increased life years and longer time spent in the alive health states, in particular KCCQ-TSS Q4 (+0.19 QALYs). The reduced incidence of HHF also contributed to QALY gains (+0.05 QALYs). The clinical outcomes of the model and disaggregated results of the base-case analysis are presented in Appendix J.

Table 54. Base-case analysis: deterministic results for empagliflozin as an add-on to standard care

Technology	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
SoC	£16,887	5.62	3.55	-	-	-	-
Empagliflozin + SoC	£17,950	5.83	3.78	£1,063	0.21	0.22	£4,804

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; SoC, standard of care; QALYs, quality-adjusted life year

B.3.8 Sensitivity analyses

B.3.8.1 Probabilistic sensitivity analysis

Probabilistic sensitivity analysis (PSA) was undertaken to translate uncertainty in the model parameters to decision uncertainty through simultaneous sampling of critical parameters from their respective distributions. The PSA encompassed parameters which inform the calculated rates of HHF, all-cause and CV mortality, as well as those informing the estimates of health state utilities, unit costs, AE and HHF utility decrements, and transition probabilities. The observed standard error was used to determine the probabilistic distribution of all parameters except costs and transition probabilities, where the standard error was calculated as a proportion of the mean value. For the costs, the standard error was assumed equal to 20% of the mean. For KCCQ-CSS transition probabilities however the standard error was assumed equal to 10% of the mean value to avoid iterations where the observed trend in the transition probabilities with empagliflozin + SoC versus SoC was inconsistent with the deterministic analyses, i.e. to ensure that relative probability of a given transition remains the same as in deterministic analysis across all probabilistic iterations.

Covariance matrices for parameters informing the rate of CV death, all-cause death, HHF, and baseline utility estimates were included in the model. Using these, a Cholesky decomposition was performed and the resulting lower-triangular matrix was then used to jointly draw samples of these parameters from a normal distribution. All cost parameters were assigned a gamma distribution, while disutilities associated with AEs and HHF were assigned the beta distribution. Details on the parameters, SEs, and assumptions are provided in Section B.3.6.1. One thousand PSA iterations were run to ensure that stable estimates of the required model outputs were obtained (see ICER convergence in Figure 25. The mean discounted total costs and mean discounted total QALYs were calculated to estimate the probabilistic ICER.

Figure 25. ICER convergence: base-case analysis

Key: ICER, Incremental cost-effectiveness ratio

Results of the PSA are summarised in the cost-effectiveness scatterplot (

Figure *26*). Each point on the chart represents a single probabilistic iteration of the model. Of one thousand iterations, 79% produced ICERs that fell below a willingness-to-pay threshold of £20,000 per QALY which is represented by the dotted line

Figure *26*. The cost-effectiveness acceptability curve in **Error! Reference source not found.** illustrates the probability of empagliflozin + SoC being cost-effective at different willingness-to-pay thresholds. At a willingness-to-pay threshold of £4,500 per QALY, empagliflozin + SoC reaches a 43% probability of being cost-effective. Conversely, at £30,000 per QALY, the probability of empagliflozin + SoC being cost-effective increases was 81%. The ICER from the PSA converged at £4,894 (<£10k per QALY) (Table 55) which was comparable to the deterministic ICER of £4,804/QALY (Table 56). The similarity between the deterministic and the probabilistic ICERs indicated that the model is sufficiently linear.

Table 55. Base-case analysis: probabilistic results for empagliflozin as an add-on to standard care

Technology	Total costs (£), mean	Total LYG, mean	Total QALYs, mean	Incremental costs (£), mean	Incremental LYG, mean	Incremental QALYs - mean	ICER incremental (£/QALY)
SoC	£16,830	5.56	3.52	-	-	-	-
Empagliflozin + SoC	£17,876	5.76	3.74	£1046	0.20	0.21	£4,894

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years; SoC, standard of care.

Figure 26. Base-case analysis: cost-effectiveness scatterplot

[redacted]

Key: QALY, Quality-adjusted Life Years

Figure 27. Base-case analysis: cost-effectiveness acceptability curve

[redacted]

Key: CEAC, Cost-effectiveness acceptability curve

B.3.8.2 Deterministic sensitivity analysis

A deterministic (or one-way) sensitivity analysis was performed to assess how changes in one specified parameter at a time impact the predicted costs and outcomes of empagliflozin + SoC compared to SoC alone, and to identify the main drivers of cost-effectiveness in the model. The most influential parameter was the treatment effect of empagliflozin + SoC associated with HHF. When this parameter was set to zero, the ICER increased to a high of £9,614//QALY (Table 56). Other drivers of ICER included the discount rates for cost and health outcomes, and the treatment effect of empagliflozin + SoC associated with all-cause mortality.

Table 56. Deterministic sensitivity analyses inputs and results

Scenario	Base-Case Input	Alternative Input	Description	ICER per QALY
Base-case	-	-	-	
Clinical Inputs				
CV & all-cause death: Distribution	Weibull	Exponential	Alternative distribution	£4,291
CV death: Treatment effect	-0.05890107	0	No treatment effect	£5,147
All-cause death: Adjust with UK lifetable?	Yes	No	No lifetable adjustment	£4,804
All-cause death: Treatment effect	-0.04435778	0	No treatment effect	£4,672
HHF: Treatment effect	-0.3245774	0	No treatment effect	£9,614
Discontinuation: Distribution	Exponential	Weibull	Alternative distribution	£4,876
Discontinuation: Treatment Effect	-0.0902196	0	No treatment effect	£4,784
Include discontinuation?	Yes	No	No discontinuation of empagliflozin	£5,037
HR for empagliflozin + SoC composite renal endpoint	0.5	0.32	Lower bound of the 95% CI	£4,149
		0.77	Upper bound of the 95% CI	£5,787
Costs and Resource Use				
Cost of HHF	£3,072	£2,426.27	Decrease by 20%	£5,328

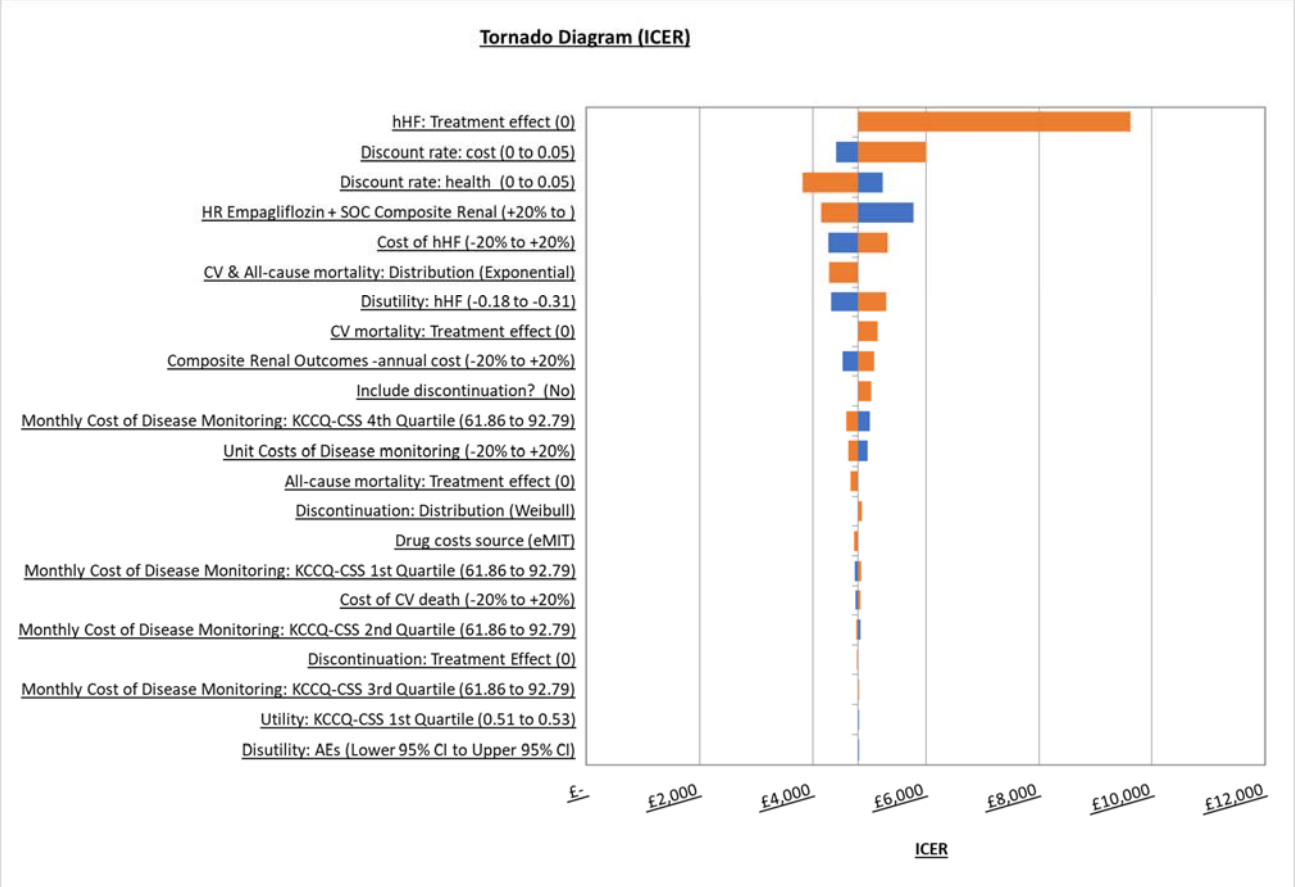
Scenario	Base-Case Input	Alternative Input	Description	ICER per QALY
		£3,639.40	Increase by 20%	£4,280
Cost of CV death	£4,146	£3,316.80	Decrease by 20%	£4,849
		£4,975.20	Increase by 20%	£4,760
Unit Costs of Disease monitoring	Multiple Values	Multiple Values	Decrease by 20%	£4,632
			Increase by 20%	£4,977
Monthly Cost of Disease Monitoring: KCCQ-CSS 1st Quartile	£77	£62	Decrease by 20%	£4,863
		£93	Increase by 20%	£4,746
Monthly Cost of Disease Monitoring: KCCQ-CSS 2nd Quartile	£77	£62	Decrease by 20%	£4,767
		£93	Increase by 20%	£6,347
Monthly Cost of Disease Monitoring: KCCQ-CSS 3rd Quartile	£77	£62	Decrease by 20%	£4,817
		£93	Increase by 20%	£4,791
Monthly Cost of Disease Monitoring: KCCQ-CSS 4th Quartile	£77	£62	Decrease by 20%	£4,598
		£93	Increase by 20%	£4,816
Cost of AE management	Multiple Values	Multiple Values	Decrease by 20%	£4,810
			Increase by 20%	£4,799
Drug costs source	MIMS	eMIT	Alternate data source	£4,736
Composite renal outcomes – annual cost	£4,862	£3,890	Decrease by 20%	£5,083
		£5,834	Increase by 20%	£4,525
Utilities				
Utility: KCCQ-CSS 1st Quartile	0.52	0.5123	Lower 95% CI	£4,816
		0.5280	Upper 95% CI	£4,816
Utility: KCCQ-CSS 2nd Quartile	0.64	0.6311	Lower 95% CI	£4,810
		0.6428	Upper 95% CI	£4,799
Utility: KCCQ-CSS 3rd Quartile	0.71	0.7042	Lower 95% CI	£4,802
		0.7159	Upper 95% CI	£4,806
Utility: KCCQ-CSS 4th Quartile	0.77	0.7681	Lower 95% CI	£4,836
		0.7799	Upper 95% CI	£4,773

Scenario	Base-Case Input	Alternative Input	Description	ICER per QALY
Disutility: HHF	-0.25	-0.184	Lower 95% CI	£5,303
		-0.308	Upper 95% CI	£4,323
Disutility: AEs	Multiple Values	Multiple Values	Lower 95% CI	£4,797
			Upper 95% CI	£4,813
Settings				
Time horizon	Lifetime	10 years	Lower range	£4,801
		20 years	Upper range	£4,803
Discount rate: cost	3.5%	0%	Lower range	£6,011
		5%	Upper range	£4,418
Discount rate: health	3.5%	0%	Lower range	£3,819
		5%	Upper range	£5,239

Abbreviations: AEs, adverse events; CV, cardiovascular; HR, hazard ratio; HHF, hospitalisation for heart failure; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire Clinical Summary Score; UK, United Kingdom.

A tornado diagram showing the impact of model parameters on the ICER is provided in Figure 28. In conclusion, the ICER for empagliflozin + SoC relative to SoC alone remained below £9,614/QALY across all parameter variations, with most analyses resulting in ICERs below £6,000/QALY.

Figure 28.Tornado diagram



Key: ICER, Incremental cost-effectiveness ratio

B.3.8.3 Scenario analysis

Scenarios analyses were run to test the validity of the structural assumptions, including the choice of parametric models to inform mortality and time-to-treatment discontinuation extrapolations, and the impact of utility age-adjustment. Table 57 provides a description of each scenario along with its resulting ICER. The results indicate that ICER is not significantly affected by change in the number of inflection points in the model, choice of parametric distribution for mortality or treatment discontinuation, utility age-adjustment, or cost associated with non-CV death.

The EMPEROR-Reduced trial was not powered to show a statistically significant difference in CV mortality or all-cause mortality between empagliflozin and placebo as add-ons to SoC (see Section B.2.8). To estimate the true effect of SGLT2 inhibition on CV and all-cause mortality, results from EMPEROR-Reduced and DAPA-HF trials

were pooled in a meta-analysis, which enabled creation of a larger sample size and derivation of treatment effect estimates with reduced variability (109). When such estimates of HRs for CV mortality, all-cause mortality, and risk of HHF with SGLT2 inhibitors vs SoC are used in the model, the resulting deterministic and probabilistic ICER is similar to that of the base-case. The deterministic and probabilistic ICER for the base case of empagliflozin + SoC vs SoC alone was £4,804/QALY and £4,894/QALY respectively. The ICER was similar for SGLT2i + SoC vs SoC; it was £4,896/QALY (deterministic) and £5,217/QALY (probabilistic). The pooled treatment effect likely represents the true treatment effect for both empagliflozin and dapagliflozin in HFrEF, since neither trial was sufficiently powered to assess effects on CV mortality or all-cause mortality.

Table 57. Scenario Analyses: ICERs for empagliflozin as add-on to standard care compared to standard care alone

Scenario	Description	ICER (Cost in £ / QALY)	% Change Relative to Base-Case ICER
Base-case		4,804	-
SGLT2i class effect	Use of the hazard ratios from Zannad et al (119) to inform all-cause mortality, CV mortality, and HHF with SGLT2i vs SoC based on pooled data from DAPA-HF and EMPEROR-Reduced.	£4,896	1.3%
One Inflection Point	Use the KCCQ quartile transition matrix used for months 4 – 8 in the model base-case from month 4 to the end of the time horizon.	5,548	15.5%
Mortality: log normal	Extrapolate CV and all-cause mortality outcomes using a log normal distribution.	3,422	-28.8%
Mortality: log-logistic	Extrapolate CV and all-cause mortality outcomes using a log-logistic distribution.	4,152	-13.6%
Mortality: Exponential	Extrapolate CV and all-cause mortality outcomes using an exponential distribution.	4,291	-10.7%
Mortality: Generalised Gamma	Extrapolate CV and all-cause mortality outcomes using a generalised gamma distribution.	4,837	0.7%
Mortality: Gompertz	Extrapolate CV and all-cause mortality outcomes using a Gompertz distribution.	5,362	11.6%
Discontinuation: Weibull	Extrapolate time to discontinuation for empagliflozin using a Weibull distribution.	4,876	1.5%
Discontinuation: log normal	Extrapolate time to discontinuation for empagliflozin using a log normal distribution.	4,906	2.1%
Discontinuation: log-logistic	Extrapolate time to discontinuation for empagliflozin using a log-logistic distribution.	4,887	1.7%
Discontinuation: Generalised Gamma	Extrapolate time to discontinuation for empagliflozin using a generalised gamma distribution.	4,847	0.9%
Discontinuation: Gompertz	Extrapolate time to discontinuation for empagliflozin using a Gompertz distribution.	4,955	3.1%
Utility: age-adjustment off	Use utility data as collected in the trial (KCCQ 4: 0.8581; KCCQ 3: 0.7942; KCCQ 2: 0.7211; KCCQ 1: 0.6043), without adjusting KCCQ 4 to be equal to UK general population utility.	4,456	-7.2%
Non-CV death costs	Assuming that non-CV deaths incur the same costs as CV deaths.	4,895	1.9%

No composite renal outcome costs and benefits	Excluding the costs and benefits of the composite renal outcome.	6,305	31.2%

Abbreviations: SoC, standard of care; SGLT2i, sodium-glucose co-transporter-2 inhibitors.
The value in the parentheses for the SGLT2i scenario corresponds to the probabilistic ICER. All the remaining ICER estimates are deterministic

B.3.8.4 Summary of sensitivity analyses results

The overall agreement between deterministic and probabilistic results suggests that the economic model is adequately linear. Furthermore, the results of the deterministic and probabilistic sensitivity analyses demonstrate that ICER is robust with respect to changes in model inputs (Section B.3.8.2 and B.3.8, respectively). In the DSA, the most influential parameter was the treatment effect of empagliflozin + SoC on the risk of HHF. The assumption of no beneficial effect of empagliflozin + SoC on the reduction of risk of HHF relative to SoC increases the ICER to £9,614 per QALY compared to the base-case value of £4,804 per QALY. The exclusion of the costs and benefits of the composite renal outcome also increased ICER by 31% to £6,305 per QALY. Other model parameters with a significant, albeit lower, impact on ICER were the discount rates for costs and health outcomes and the effect of the treatment on all-cause mortality. Scenario analyses validated assumptions underlying the model structure since the choice of parametric distribution for all-cause mortality, CV mortality or treatment discontinuation did not considerably alter the estimates of the ICER. Similarly, the number of inflection points or the utility age adjustment did not have a sizable impact on the model outcomes. The results of the SGLT2i analysis also demonstrated the cost-effectiveness of the SGLT2i class against SoC. Across all sensitivity and scenario analyses ICER estimates remained well below the usual willingness-to-pay thresholds.

B.3.9 Subgroup analysis

As mentioned in Section B.2.7, all patient subgroups stand to benefit from empagliflozin as an add-on to SoC regardless of whether comorbidities are present or not. In the EMPEROR-Reduced trial a reduction in HHF or adjudicated CV death (composite primary outcome) was shown across multiple subgroups, including age (<65yr/>65yr), sex (male/female), race (White, Black, Asian, other), body-mass index, and prior therapies (ARNI/no ARNI). Thus, only the ITT population was considered in the economic analysis.

B.3.10 Validation

Prior to submission, the cost-effectiveness model was quality-assured through internal processes and by external economists. An economist who had not been involved in the model development reviewed the model for coding errors, inconsistencies, and the plausibility of inputs. This was done as a thorough sheet-by-sheet and cell-by-cell check. The model was also reviewed against a checklist of known modelling errors and questioning of assumptions through white-box and black-box tests; the checklist followed was based on publicly available and peer-reviewed checklists (200). More information on the model verification process is available in the IQVIA validation report (201).

B.3.10.1 Validation of cost-effectiveness analysis

B.3.10.1.1 Internal validity

Internal validation was undertaken to assess the model's ability to accurately predict the observed outcomes from the EMPEROR-Reduced trial. The rates of HHF, CV death, and all-cause death observed during the trial follow-up of 16 months were compared with the economic model predictions over an 18-month time horizon. Figure 29 shows the observed and predicted CV death, all-cause death, and HHF rates per 100 patient years, respectively, obtained with the simple risk equations containing only treatment and time-varying KCCQ-CSS health states as predictors as well as those including the full set of predictors in the ITT population. It is evident that the economic model that used simpler risk equations accurately predicted the observed rates across all outcomes and subgroups. By contrast, the equations with the full set of predictor variables consistently underpredicted the event rates.

Figure 29. Comparison of observed and model-predicted event rates in the ITT population: CV death, all-cause death, and HHF rates per 100 patient years

B.3.10.1.2 External validity

The external validity of the economic model predictions was checked against the predictions of the dapagliflozin model described in TA697 (6) and the PULSE study (43). As shown in Appendix O, the long-term predictions of all-cause mortality and the estimated all-cause death rates were closely aligned with predictions from the dapagliflozin model in TA697 and with observed rates of all-cause death in the PULSE study, respectively.

However, some discrepancies in HHF and CV death were noted. In the comparison against TA697, significantly higher rates of HHF predicted by empagliflozin model were attributed to the differences in the EMPEROR-Reduced and DAPA-HF trial populations and the fact that EMPEROR-Reduced was enriched for patients with a more severe disease who were at higher risk of HHF. Also, HHF and CV death rates predicted by the empagliflozin model were notably higher than those observed in the PULSE study. Again, this was likely due to the enriched EMPEROR-Reduced population having more severe disease and a higher risk of adverse CV outcomes compared to the CPRD cohort. Further, since PULSE relied on ONS mortality data for primary cause of death, the rate of CV death is likely to be under-recorded compared to EMPEROR-Reduced, where CV death was an adjudicated endpoint. Further information on the external validation can be found in Appendix O and the IQVIA validation report (201).

B.3.11 Interpretation and conclusions of economic evidence

The cost-effectiveness model for economic evaluation of empagliflozin + SoC builds on the modelling approach previously accepted by the NICE committee for TA697 (6). Model inputs were primarily derived from the EMPEROR-Reduced trial, including inputs for baseline characteristics, health state transition probabilities, health state utility values, disutilities associated with clinical events, survival equations, risk equations, AE incidence rates, and treatment discontinuation rates. Additional model inputs for disutilities of AEs, unit costs, and resource use were identified from the published literature or from NHS National Reference Costs. The model was able to reproduce the EMPEROR-Reduced trial results over the mean trial follow-up period of 16 months and was used to extrapolate those results to a lifetime horizon.

In the base-case analysis, over the lifetime horizon, patients treated with empagliflozin + SoC experienced a lower rate of HHF (17.1 per 100 PY vs 20.80 per 100 PY on SoC) and CV death (9.81 per 100 PY vs 10.34 per 100 PY on SoC) compared to those treated with SoC alone. The difference in the rate of non-CV death between empagliflozin + SoC and SoC arm was minimal (4.41 per 100 PY vs. 4.46 per 100 PY). Reduction in clinical event rates with empagliflozin + SoC compared to SoC was the key driver of the incremental benefits, while incremental costs were largely attributable to empagliflozin + SoC drug costs which were in part due to the longer survival and treatment duration of patients receiving empagliflozin as an add-on to SoC (average time of receiving empagliflozin + SoC was 4.04 years). The base-case analysis estimated a probabilistic (deterministic) ICER of £4,894 (£4,804) (< £10k) per QALY gained suggesting that empagliflozin + SoC offers a good use of NHS resources and should be preferred over SoC alone based on usual threshold values.

Sensitivity and scenario analyses showed that the cost-effectiveness model was robust to variation in model parameters. The probabilistic and deterministic base-case results were closely aligned, with [redacted] of the iterations falling in the north-east quadrant of the cost-effectiveness plane suggesting a high probability of empagliflozin + SoC being more costly and more effective than SoC alone. The deterministic sensitivity

analyses and scenario analyses were associated with ICERs of less than £9,614 per QALY gained with most of scenarios yielding ICERs below £6,000 per QALY gained. The most influential parameter identified in one-way sensitivity analysis was the treatment effect of empagliflozin + SoC on the rate of HHF.

The results of the analysis should be interpreted considering its limitations. Firstly, the EMPEROR-Reduced trial offered only short-term data and therefore long-term outcomes had to be extrapolated at the expense of uncertainty. Although this limitation is inherent to most cost-effectiveness models, its impact on the overall decision uncertainty has been mitigated through comparison of model predictions with those approved by the NICE committee in TA697 (6, 201). Also, sensitivity analyses indicated that the choice of the parametric model did not have a significant impact on the estimated ICER. Secondly, the economic analysis assumes that the HF event rates observed in UK clinical practice mirror those observed in the EMPEROR-Reduced trial. The relevance of the trial to the UK clinical practice is strengthened by the protocol requirement for patients to receive stable doses of guideline-recommended HF therapies at baseline. Thirdly, the model does not capture diabetic ketoacidosis, a rare but recognised complication of SGLT2 inhibition, because no such cases were observed in the EMPEROR-Reduced trial. This assumption is further supported by the EMPA-REG OUTCOME trial in patients with T2D and established CV disease, where few cases were observed and there was no imbalance between treatment groups (163).

In conclusion, the cost-effectiveness analysis presented here demonstrates that empagliflozin represents a cost-effective use of NHS resources as an add-on to standard care therapy for the treatment of HFrEF. The decision analytic model underlying the economic analysis is closely aligned with the model presented in TA697 which evaluated dapagliflozin as an add-on to standard care for HFrEF (6).

B.4 References

1. European Medicines Agency. Jardiance 2014 [Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/jardiance.
2. National Institute for Health and Care Excellence. Canagliflozin, dapagliflozin and empagliflozin as monotherapies for treating type 2 diabetes Technology appraisal guidance [TA390] 2016 [Available from: https://www.nice.org.uk/guidance/ta390.
3. National Institute for Health and Care Excellence. Empagliflozin in combination therapy for treating type 2 diabetes Technology appraisal guidance [TA336] 2015 [Available from: https://www.nice.org.uk/guidance/ta336.
4. National Institute for Health and Care Excellence. Final appraisal document: Dapagliflozin for treating chronic heart failure with reduced ejection fraction. 2020.
5. IQVIA. LPD, IQVIA Ltd, incorporating data derived from THIN, A Cegedim Database, May 2021.
6. NICE. Dapagliflozin for treating chronic heart failure with reduced ejection fraction [TA 679] 2021 [Available from: https://www.nice.org.uk/guidance/ta679.
7. NICE. Technology appraisal guidance [TA398]: Lumacaftor–ivacaftor for treating cystic fibrosis homozygous for the F508del mutation 2016 [Available from: https://www.nice.org.uk/guidance/ta398/chapter/committee-discussion#cost-effectiveness-2.
8. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. Process and methods [PMG9]. 2013 [Available from: https://www.nice.org.uk/process/pmg9/chapter/the-reference-case.
9. Conrad N, Judge A, Tran J, Mohseni H, Hedgecott D, Crespillo AP, et al. Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. The Lancet. 2018;391(10120):572-80.
10. Lawson Claire A, Zaccardi F, Squire I, Okhai H, Davies M, Huang W, et al. Risk Factors for Heart Failure. Circulation: Heart Failure. 2020;13(2):e006472.
11. Moscelli G, Siciliani L, Gutacker N, Cookson R. Socioeconomic inequality of access to healthcare: Does choice explain the gradient? J Health Econ. 2018;57:290-314.
12. McCartney G, Hart C, Watt G. How can socioeconomic inequalities in hospital admissions be explained? A cohort study. BMJ Open. 2013;3(8):e002433.
13. National Institute for Health and Care Excellence. The NICE strategy 2021 to 2026 2021 [Available from: https://www.nice.org.uk/about/who-we-are/corporate-publications/the-nice-strategy-2021-to-2026.
14. NICE. Our Priniciples: Social Value Judgements 2021 [Available from: https://www.nice.org.uk/about/who-we-are/our-principles.
15. Foundation H. Build Back Fairer: The COVID-19 Marmot Review. 2020.
16. Ndefo UA, Anidiobi NO, Basheer E, Eaton AT. Empagliflozin (Jardiance): A Novel SGLT2 Inhibitor for the Treatment of Type-2 Diabetes. P T. 2015;40(6):364-8.
17. Lopaschuk GD, Verma S. Mechanisms of Cardiovascular Benefits of Sodium Glucose Co-Transporter 2 (SGLT2) Inhibitors: A State-of-the-Art Review. JACC Basic Transl Sci. 2020;5(6):632-44.
18. Cowie MR, Fisher M. SGLT2 inhibitors: mechanisms of cardiovascular benefit beyond glycaemic control. Nature Reviews Cardiology. 2020;17(12):761-72.

19. Summary of Product Characteristics. Jardiance 2020 [Available from: https://www.ema.europa.eu/en/documents/product-information/jardiance-epar-product-information_en.pdf.
20. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur J Heart Fail. 2016;18(8):891-975.
21. Murphy SP, Ibrahim NE, Januzzi JL, Jr. Heart Failure With Reduced Ejection Fraction: A Review. JAMA. 2020;324(5):488-504.
22. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Jr., Drazner MH, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol. 2013;62(16):e147-239.
23. National Institute for Health and Care Excellence. Resource impact report: Chronic heart failure in adults: diagnosis and management (NG106). 2018.
24. Ziaeian B, Fonarow GC. Epidemiology and aetiology of heart failure. Nat Rev Cardiol. 2016;13(6):368-78.
25. Kenchaiah S, Evans JC, Levy D, Wilson PWF, Benjamin EJ, Larson MG, et al. Obesity and the Risk of Heart Failure. New England Journal of Medicine. 2002;347(5):305-13.
26. Matthews JC, Dardas TF, Dorsch MP, Aaronson KD. Right-sided heart failure: diagnosis and treatment strategies. Curr Treat Options Cardiovasc Med. 2008;10(4):329-41.
27. Mangini S, Pires PV, Braga FG, Bacal F. Decompensated heart failure. Einstein (Sao Paulo). 2013;11(3):383-91.
28. Nhfa Csanz Heart Failure Guidelines Working Group, Atherton JJ, Sindone A, De Pasquale CG, Driscoll A, MacDonald PS, et al. National Heart Foundation of Australia and Cardiac Society of Australia and New Zealand: Guidelines for the Prevention, Detection, and Management of Heart Failure in Australia 2018. Heart Lung Circ. 2018;27(10):1123-208.
29. Voelkel NF, Quaife RA, Leinwand LA, Barst RJ, McGoon MD, Meldrum DR, et al. Right Ventricular Function and Failure. Circulation. 2006;114(17):1883-91.
30. Butler J, Fonarow GC, Zile MR, Lam CS, Roessig L, Schelbert EB, et al. Developing therapies for heart failure with preserved ejection fraction: current state and future directions. JACC Heart Fail. 2014;2(2):97-112.
31. Lam CS, Teng TH. Understanding Heart Failure With Mid-Range Ejection Fraction. JACC Heart Fail. 2016;4(6):473-6.
32. Raphael C, Briscoe C, Davies J, Ian Whinnett Z, Manisty C, Sutton R, et al. Limitations of the New York Heart Association functional classification system and self-reported walking distances in chronic heart failure. Heart. 2007;93(4):476-82.
33. British Heart Foundation. Heart statistics: UK factsheet 2021 [Available from: https://www.bhf.org.uk/what-we-do/our-research/heart-statistics.
34. National Institute for Health and Care Excellence. Chronic heart failure in adults: diagnosis and management NICE guideline [NG106] 2018 [Available from: https://www.nice.org.uk/guidance/ng106.

35. Boehringer Ingelheim. Understanding the heart failure patient journey [data on file]. 2020.
36. Groenewegen A, Rutten FH, Mosterd A, Hoes AW. Epidemiology of heart failure. Eur J Heart Fail. 2020;22(8):1342-56.
37. Mamas MA, Sperrin M, Watson MC, Coutts A, Wilde K, Burton C, et al. Do patients have worse outcomes in heart failure than in cancer? A primary care-based cohort study with 10-year follow-up in Scotland. Eur J Heart Fail. 2017;19(9):1095-104.
38. Morton G, Masters J, Cowburn PJ. Multidisciplinary team approach to heart failure management. Heart. 2018;104(16):1376-82.
39. Masters H, Freeman J, Dixon S. Variable structure and provision of guideline-based care in specialist heart failure centres in the UK: a survey of 100 health professionals. British Journal of Cardiac Nursing. 2020;15(5).
40. NHS_Digital. Quality and Outcomes Framework, 2019-20: Prevalence, achievement and personalised care adjustments at regional and national level 2020 2019-2020 [Available from: https://digital.nhs.uk/data-and-information/publications/statistical/quality-and-outcomes-framework-achievement-prevalence-and-exceptions-data/2019-20.
41. National Heart Failure Audit. Summary Report (2017/18 data). 2019.
42. Dulai R, Sheikh AS, Qureshi A, Katechia S, Peysakhova Y, Johns M, et al. Prevalence, clinical characteristics and outcomes of HF with preserved versus reduced ejection fraction. The British Journal of Cardiology. 2016;23(1).
43. Boehringer Ingelheim. Incidence prevalence and resource utilisation of heart failure in England: a linked descriptive analysis of English primary and secondary care – PULSE study. [Data on File]. 2021.
44. British Lung Foundation. Chronic obstructive pulmonary disease (COPD) statistics 2020 [Available from: https://statistics.blf.org.uk/copd.
45. Cancer Research UK. Cancer incidence statistics 2017 [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence#heading-One.
46. Rangaswami J, Bhalla V, Blair JEA, Chang TI, Costa S, Lentine KL, et al. Cardiorenal Syndrome: Classification, Pathophysiology, Diagnosis, and Treatment Strategies: A Scientific Statement From the American Heart Association. Circulation. 2019;139(16):e840-e78.
47. Dhingra R, Vasan RS. Diabetes and the risk of heart failure. Heart Fail Clin. 2012;8(1):125-33.
48. Cherney DZI, Repetto E, Wheeler DC, Arnold SV, MacLachlan S, Hunt PR, et al. Impact of Cardio-Renal-Metabolic Comorbidities on Cardiovascular Outcomes and Mortality in Type 2 Diabetes Mellitus. Am J Nephrol. 2020;51(1):74-82.
49. Lofman I, Szummer K, Hagerman I, Dahlstrom U, Lund LH, Jernberg T. Prevalence and prognostic impact of kidney disease on heart failure patients. Open Heart. 2016;3(1):e000324.
50. House AA, Wanner C, Sarnak MJ, Piña IL, McIntyre CW, Komenda P, et al. Heart failure in chronic kidney disease: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference. Kidney Int. 2019;95(6):1304-17.

51. Lawson CA, Jones PW, Teece L, Dunbar SB, Seferovic PM, Khunti K, et al. Association Between Type 2 Diabetes and All-Cause Hospitalization and Mortality in the UK General Heart Failure Population: Stratification by Diabetic Glycemic Control and Medication Intensification. JACC Heart Fail. 2018;6(1):18-26.
52. Kenny HC, Abel ED. Heart Failure in Type 2 Diabetes Mellitus. Circ Res. 2019;124(1):121-41.
53. Yang H, Negishi K, Otahal P, Marwick TH. Clinical prediction of incident heart failure risk: a systematic review and meta-analysis. Open Heart. 2015;2(1):e000222.
54. Hamo C, Kwak L, Florido R, Echouffo-Tcheugui J, Blumenthal R, Loehr LR, et al. Abstract 14768: Modifiable Risk Factor Control and Risk for Heart Failure With Preserved Ejection Fraction (HFpEF) versus Reduced Ejection Fraction (HFrEF): The Atherosclerosis Risk in Communities Study (ARIC). Circulation. 2019;140(1).
55. Kubicki DM, Xu M, Akwo EA, Dixon D, Munoz D, Blot WJ, et al. Race and Sex Differences in Modifiable Risk Factors and Incident Heart Failure. JACC Heart Fail. 2020;8(2):122-30.
56. Zhang J, Hobkirk J, Carroll S, Pellicori P, Clark AL, Cleland JG. Exploring quality of life in patients with and without heart failure. Int J Cardiol. 2016;202:676-84.
57. Taylor CJ, Ordóñez-Mena JM, Roalfe AK, Lay-Flurrie S, Jones NR, Marshall T, et al. Trends in survival after a diagnosis of heart failure in the United Kingdom 2000-2017: population based cohort study. BMJ. 2019;364:l223.
58. Taylor CJ, Ryan R, Nichols L, Gale N, Hobbs FR, Marshall T. Survival following a diagnosis of heart failure in primary care. Family Practice. 2017;34(2):161-8.
59. Drozd M, Relton SD, Walker AMN, Slater TA, Gierula J, Paton MF, et al. Association of heart failure and its comorbidities with loss of life expectancy. Heart. 2020:heartjnl-2020-317833.
60. Aguilar D. Heart Failure, Diabetes Mellitus, and Chronic Kidney Disease: A Clinical Conundrum. Circulation: Heart Failure. 2016;9(7).
61. Ouwerkerk W, Voors AA, Anker SD, Cleland JG, Dickstein K, Filippatos G, et al. Determinants and clinical outcome of uptitration of ACE-inhibitors and beta-blockers in patients with heart failure: a prospective European study. Eur Heart J. 2017;38(24):1883-90.
62. Komajda M, Follath F, Swedberg K, Cleland J, Aguilar JC, Cohen-Solal A, et al. The EuroHeart Failure Survey programme--a survey on the quality of care among patients with heart failure in Europe. Part 2: treatment. Eur Heart J. 2003;24(5):464-74.
63. Crespo-Leiro MG, Anker SD, Maggioni AP, Coats AJ, Filippatos G, Ruschitzka F, et al. European Society of Cardiology Heart Failure Long-Term Registry (ESC-HF-LT): 1-year follow-up outcomes and differences across regions. Eur J Heart Fail. 2016;18(6):613-25.
64. Waldum B, Westheim AS, Sandvik L, Flonaes B, Grundtvig M, Gullestad L, et al. Renal function in outpatients with chronic heart failure. J Card Fail. 2010;16(5):374-80.
65. Tromp J, Bamadhaj S, Cleland JGF, Angermann CE, Dahlstrom U, Ouwerkerk W, et al. Post-discharge prognosis of patients admitted to hospital for

heart failure by world region, and national level of income and income disparity (REPORT-HF): a cohort study. Lancet Glob Health. 2020;8(3):e411-e22.
66. Dauriz M, Targher G, Temporelli PL, Lucci D, Gonzini L, Nicolosi GL, et al. Prognostic Impact of Diabetes and Prediabetes on Survival Outcomes in Patients With Chronic Heart Failure: A Post-Hoc Analysis of the GISSI-HF (Gruppo Italiano per lo Studio della Sopravvivenza nella Insufficienza Cardiaca-Heart Failure) Trial. J Am Heart Assoc. 2017;6(7).
67. Kristensen SL, Mogensen UM, Jhund PS, Petrie MC, Preiss D, Win S, et al. Clinical and Echocardiographic Characteristics and Cardiovascular Outcomes According to Diabetes Status in Patients With Heart Failure and Preserved Ejection Fraction: A Report From the I-Preserve Trial (Irbesartan in Heart Failure With Preserved Ejection Fraction). Circulation. 2017;135(8):724-35.
68. Lawson CA, Testani JM, Mamas M, Damman K, Jones PW, Teece L, et al. Chronic kidney disease, worsening renal function and outcomes in a heart failure community setting: A UK national study. International journal of cardiology. 2018;267:120-7.
69. van Riet EE, Hoes AW, Wagenaar KP, Limburg A, Landman MA, Rutten FH. Epidemiology of heart failure: the prevalence of heart failure and ventricular dysfunction in older adults over time. A systematic review. Eur J Heart Fail. 2016;18(3):242-52.
70. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, et al. Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. Circulation. 2019;139(10):e56-e528.
71. Boonman-de Winter LJ, Rutten FH, Cramer MJ, Landman MJ, Liem AH, Rutten GE, et al. High prevalence of previously unknown heart failure and left ventricular dysfunction in patients with type 2 diabetes. Diabetologia. 2012;55(8):2154-62.
72. Lin AH, Chin JC, Sicignano NM, Evans AM. Repeat Hospitalizations Predict Mortality in Patients With Heart Failure. Mil Med. 2017;182(9):e1932-e7.
73. Thokala P, Baalbaki H, Brennan A, Pandor A, Stevens JW, Gomersall T, et al. Telemonitoring after discharge from hospital with heart failure: cost-effectiveness modelling of alternative service designs. BMJ Open. 2013;3(9):e003250.
74. Murcia AM, Hennekens CH, Lamas GA, Jimenez-Navarro M, Rouleau JL, Flaker GC, et al. Impact of diabetes on mortality in patients with myocardial infarction and left ventricular dysfunction. Arch Intern Med. 2004;164(20):2273-9.
75. MacDonald MR, Petrie MC, Varyani F, Ostergren J, Michelson EL, Young JB, et al. Impact of diabetes on outcomes in patients with low and preserved ejection fraction heart failure: an analysis of the Candesartan in Heart failure: Assessment of Reduction in Mortality and morbidity (CHARM) programme. Eur Heart J. 2008;29(11):1377-85.
76. Shah AM, Uno H, Kober L, Velazquez EJ, Maggioni AP, MacDonald MR, et al. The inter-relationship of diabetes and left ventricular systolic function on outcome after high-risk myocardial infarction. Eur J Heart Fail. 2010;12(11):1229-37.
77. Ter Maaten JM, Martens P, Damman K, Dickstein K, Ponikowski P, Lang CC, et al. Higher doses of loop diuretics limit uptitration of angiotensin-converting enzyme inhibitors in patients with heart failure and reduced ejection fraction. Clin Res Cardiol. 2020;109(8):1048-59.

78. Cook C, Cole G, Asaria P, Jabbour R, Francis DP. The annual global economic burden of heart failure. Int J Cardiol. 2014;171(3):368-76.
79. OECD. Exchange rates (indicator) 2020 [Available from: https://data.oecd.org/conversion/exchange-rates.htm.
80. Cowie MR. The heart failure epidemic: a UK perspective. Echo Res Pract. 2017;4(1):R15-R20.
81. Olchanski N, Vest AR, Cohen JT, DeNofrio D. Comparing inpatient costs of heart failure admissions for patients with reduced and preserved ejection fraction with or without type 2 diabetes. Cardiovasc Endocrinol Metab. 2020;9(1):17-23.
82. Du H, Farmer R, Ustyugova A. Non-elective in-patient burden attributable to Cardio-Renal-Metabolic (CRM) conditions in England. [Data on File]. 2021.
83. Farré N, Vela E, Clèries M, Bustins M, Cainzos-Achirica M, Enjuanes C, et al. Real world heart failure epidemiology and outcome: A population-based analysis of 88,195 patients. PLoS One. 2017;12(2):e0172745-e.
84. Au AG, McAlister FA, Bakal JA, Ezekowitz J, Kaul P, van Walraven C. Predicting the risk of unplanned readmission or death within 30 days of discharge after a heart failure hospitalization. Am Heart J. 2012;164(3):365-72.
85. Hollingworth W, Biswas M, Maishman RL, Dayer MJ, McDonagh T, Purdy S, et al. The healthcare costs of heart failure during the last five years of life: A retrospective cohort study. International Journal of Cardiology. 2016;224:132-8.
86. Lesyuk W, Kriza C, Kolominsky-Rabas P. Cost-of-illness studies in heart failure: a systematic review 2004-2016. BMC Cardiovasc Disord. 2018;18(1):74.
87. Shafie AA, Tan YP, Ng CH. Systematic review of economic burden of heart failure. Heart Fail Rev. 2018;23(1):131-45.
88. Calado F, Viriato D, Proenca CC, Lahoz R, Agrawal R, Cotton S, et al. Burden of chronic heart failure on caregivers: a multinational cross-sectional survey in real world setting Value in Health. 2017;20(A623).
89. Berry C, McMurray J. Undertreatment of heart failure has high cost to patients. BMJ. 2001;322(7288):731-2.
90. Chen X, Xin Y, Hu W, Zhao Y, Zhang Z, Zhou Y. Quality of life and outcomes in heart failure patients with ejection fractions in different ranges. PLoS One. 2019;14(6):e0218983.
91. CV Outcomes I. The Kansas City Cardiomyopathy Questionnaire (KCCQ) 2021 [Available from: https://cvoutcomes.org/pages/3214.
92. Spertus JA, Jones PG, Sandhu AT, Arnold SV. Interpreting the Kansas City Cardiomyopathy Questionnaire in Clinical Trials and Clinical Care: JACC State-of-the-Art Review. Journal of the American College of Cardiology. 2020;76(20):2379-90.
93. O'Donnell J, Smith-Byrne K, Velardo C, Conrad N, Salimi-Khorshidi G, Doherty A, et al. Self-reported and objectively measured physical activity in people with and without chronic heart failure: UK Biobank analysis. Open Heart. 2020;7(1):e001099.
94. Zugck C, Stork S, Stockl G, investigators RE-Cs. Long-term treatment with ivabradine over 12months in patients with chronic heart failure in clinical practice: Effect on symptoms, quality of life and hospitalizations. Int J Cardiol. 2017;240:258-64.

95. Joyce E, Chung C, Badloe S, Odutayo K, Desai A, Givertz MM, et al. Variable Contribution of Heart Failure to Quality of Life in Ambulatory Heart Failure With Reduced, Better, or Preserved Ejection Fraction. JACC Heart Fail. 2016;4(3):184-93.
96. Hoekstra T, Jaarsma T, van Veldhuisen DJ, Hillege HL, Sanderman R, Lesman-Leegte I. Quality of life and survival in patients with heart failure. Eur J Heart Fail. 2013;15(1):94-102.
97. Zuluaga MC, Guallar-Castillón P, López-García E, Banegas JR, Conde-Herrera M, Olcoz-Chiva M, et al. Generic and disease-specific quality of life as a predictor of long-term mortality in heart failure. Eur J Heart Fail. 2010;12(12):1372-8.
98. Conrad N, Judge A, Canoy D, Tran J, O'Donnell J, Nazarzadeh M, et al. Diagnostic tests, drug prescriptions, and follow-up patterns after incident heart failure: A cohort study of 93,000 UK patients. PLoS Med. 2019;16(5):e1002805.
99. Hancock HC, Close H, Fuat A, Murphy JJ, Hungin AP, Mason JM. Barriers to accurate diagnosis and effective management of heart failure have not changed in the past 10 years: a qualitative study and national survey. BMJ Open. 2014;4(3):e003866.
100. Bottle A, Kim D, Aylin P, Cowie MR, Majeed A, Hayhoe B. Routes to diagnosis of heart failure: observational study using linked data in England. Heart. 2018;104(7):600.
101. Bottle A, Kim D, Aylin PP, Majeed FA, Cowie MR, Hayhoe B. Real-world presentation with heart failure in primary care: do patients selected to follow diagnostic and management guidelines have better outcomes? Open Heart. 2018;5(2):e000935.
102. Lawson CA, Zaccardi F, Squire I, Ling S, Davies MJ, Lam CSP, et al. 20-year trends in cause-specific heart failure outcomes by sex, socioeconomic status, and place of diagnosis: a population-based study. Lancet Public Health. 2019;4(8):e406-e20.
103. McMurray JP, M. How Should We Sequence the Treatments for Heart Failure and a Reduced Ejection Fraction? Circulation. 2021;143(9):875-7.
104. Zhang YB, Chen C, Pan XF, Guo J, Li Y, Franco OH, et al. Associations of healthy lifestyle and socioeconomic status with mortality and incident cardiovascular disease: two prospective cohort studies. BMJ. 2021;373:n604.
105. Sannigrahi S, Pilla F, Basu B, Basu AS, Molter A. Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. Sustain Cities Soc. 2020;62:102418.
106. Williamson E. Factors associated with COVID-19-related death using OpenSAFELY. Nature. 2020;584:430-6.
107. National Heart Failure Audit. Summary Report (2018/19 data) 2020 [Available from: https://www.nicor.org.uk/wp-content/uploads/2020/12/National-Heart-Failure-Audit-2020-FINAL.pdf.
108. British Heart Foundation. Heart failure: a blueprint for change. 2021.
109. Packer M, Anker SD, Butler J, Filippatos G, Pocock SJ, Carson P, et al. Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure. New England Journal of Medicine. 2020;383(15):1413-24.
110. Patorno E, Najafzadeh M, Pawar A, Franklin JM, Deruaz-Luyet A, Brodovicz KG, et al. The EMPagliflozin compaRative effectIveness and SafEty (EMPRISE)

study programme: Design and exposure accrual for an evaluation of empagliflozin in routine clinical care. Endocrinol Diabetes Metab. 2020;3(1):e00103.
111. Zannad F, Ferreira JP, Pocock SJ, Zeller C, Anker SD, Butler J, et al. Cardiac and Kidney Benefits of Empagliflozin in Heart Failure Across the Spectrum of Kidney Function: Insights From EMPEROR-Reduced. Circulation. 2021;143(4):310-21.
112. Packer M, Anker SD, Butler J, Filippatos G, Ferreira JP, Pocock SJ, et al. Influence of neprilysin inhibition on the efficacy and safety of empagliflozin in patients with chronic heart failure and a reduced ejection fraction: the EMPEROR-Reduced trial. Eur Heart J. 2021;42(6):671-80.
113. Anker SD, Butler J, Filippatos G, Khan MS, Marx N, Lam CSP, et al. Effect of Empagliflozin on Cardiovascular and Renal Outcomes in Patients With Heart Failure by Baseline Diabetes Status: Results From the EMPEROR-Reduced Trial. Circulation. 2021;143(4):337-49.
114. Cosentino F, Grant PJ, Aboyans V, Bailey CJ, Ceriello A, Delgado V, et al. 2019 ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD: The Task Force for diabetes, pre-diabetes, and cardiovascular diseases of the European Society of Cardiology (ESC) and the European Association for the Study of Diabetes (EASD). European Heart Journal. 2020;41(2):255-323.
115. Solomon SD, de Boer RA, DeMets D, Hernandez AF, Inzucchi SE, Kosiborod MN, et al. Dapagliflozin in Heart Failure with Preserved and Mildly Reduced Ejection Fraction: Rationale and Design of the DELIVER Trial. Eur J Heart Fail. 2021.
116. Cookson R, Propper C, Asaria M, Raine R. Socio-Economic Inequalities in Health Care in England. Fiscal Studies. 2016;37(3-4):371-403.
117. Fersia O, Bryant S, Nicholson R, McMeeken K, Brown C, Donaldson B, et al. The impact of the COVID-19 pandemic on cardiology services. Open Heart. 2020;7(2).
118. Propper C, Stoye G, Zaranko B. The Wider Impacts of the Coronavirus Pandemic on the NHS*. Fiscal Studies. 2020;41(2):345-56.
119. Zannad F, Ferreira JP, Pocock SJ, Anker SD, Butler J, Filippatos G, et al. SGLT2 inhibitors in patients with heart failure with reduced ejection fraction: a meta-analysis of the EMPEROR-Reduced and DAPA-HF trials. Lancet. 2020;396(10254):819-29.
120. Butler J, Anker SD, Filippatos G, Khan MS, Ferreira JP, Pocock SJ, et al. Empagliflozin and health-related quality of life outcomes in patients with heart failure with reduced ejection fraction: the EMPEROR-Reduced trial. Eur Heart J. 2021.
121. Packer M, Butler J, Filippatos GS, Jamal W, Salsali A, Schnee J, et al. Evaluation of the effect of sodium-glucose co-transporter 2 inhibition with empagliflozin on morbidity and mortality of patients with chronic heart failure and a reduced ejection fraction: rationale for and design of the EMPEROR-Reduced trial. Eur J Heart Fail. 2019;21(10):1270-8.
122. Zannad FF FG, Butler J, et al. Design and rationale of the Empagliflozin outcome trial in patients with chronic heart failure (EMPEROR-Reduced). Eur J Heart Fail. 2018;20(441).
123. Anker SD ZF, Butler J et al. Design and rationale of the EMPEROR trials of empagliflozin in patients with chronic heart failure with reduced or preserved ejection fraction. . Diabetologia. 2018;61(S324).

124. ClinicalTrials.gov. EMPagliflozin outcomE tRial in Patients With chrOnic heaRt Failure With Reduced Ejection Fraction (EMPEROR-Reduced) 2020 [Available from: https://clinicaltrials.gov/ct2/show/NCT03057977.
125. ClinicalTrials.gov. EMPagliflozin outcomE tRial in Patients With chrOnic heaRt Failure With Preserved Ejection Fraction (EMPEROR-Preserved) 2020 [Available from: https://clinicaltrials.gov/ct2/show/NCT03057951.
126. ClinicalTrials.gov. This Study Tests Empagliflozin in Patients With Chronic Heart Failure With Reduced Ejection Fraction (HFrEF). The Study Looks at How Far Patients Can Walk in 6 Minutes and at Their Heart Failure Symptoms 2019 [Available from: https://clinicaltrials.gov/ct2/show/NCT03448419.
127. ClinicalTrials.gov. This Study Tests Empagliflozin in Patients With Chronic Heart Failure With Preserved Ejection Fraction (HFpEF). The Study Looks at How Far Patients Can Walk in 6 Minutes and at Their Heart Failure Symptoms. 2019 [Available from: https://clinicaltrials.gov/ct2/show/NCT03448406.
128. ClinicalTrials.gov. BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients (EMPA-REG OUTCOME). 2016 [Available from: https://clinicaltrials.gov/ct2/show/NCT01131676.
129. ClinicalTrials.gov. A Study to Test the Effect of Empagliflozin in Patients Who Are in Hospital for Acute Heart Failure 2020 [Available from: https://clinicaltrials.gov/ct2/show/NCT04157751.
130. ClinicalTrials.gov. EMPA-KIDNEY (The Study of Heart and Kidney Protection With Empagliflozin) 2020 [Available from: https://clinicaltrials.gov/ct2/show/NCT03594110.
131. ClinicalTrials.gov. A Study That Looks at the Function of the Heart in Patients With Heart Failure Who Take Empagliflozin 2017 [Available from: https://clinicaltrials.gov/ct2/show/NCT03332212.
132. ClinicalTrials.gov. EMPACT-MI: A Study to Test Whether Empagliflozin Can Lower the Risk of Heart Failure and Death in People Who Had a Heart Attack (Myocardial Infarction) 2020 [Available from: https://clinicaltrials.gov/ct2/show/NCT04509674.
133. ClinicalTrials.gov. Comparative Effectiveness of Empagliflozin in the US 2020 [Available from: https://clinicaltrials.gov/ct2/show/NCT03363464.
134. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. EUPAS20677 2017 [Available from: http://www.encepp.eu/encepp/viewResource.htm?id=21657.
135. Boehringer Ingelheim International GmbH. A phase III randomised, double-blind trial to evaluate efficacy and safety of once daily empagliflozin 10 mg compared to placebo, in patients with chronic Heart Failure with reduced Ejection Fraction (HFrEF). Clinical Trial Report c28576542-012020. 2020.
136. CONSORT. CONSORT 2010 2010 [Available from: http://www.consort-statement.org/checklists/view/32--consort-2010/66-title.
137. Hicks Karen A, Tcheng James E, Bozkurt B, Chaitman Bernard R, Cutlip Donald E, Farb A, et al. 2014 ACC/AHA Key Data Elements and Definitions for Cardiovascular Endpoint Events in Clinical Trials. Journal of the American College of Cardiology. 2015;66(4):403-69.

138. Rubin DB. Multiple Imputation for Nonresponse in Surveys. . New York1987. Available from: https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/9780470316696.fmatter.
139. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association. 1999;94(446):496-509.
140. Rogers JK, Yaroshinsky A, Pocock SJ, Stokar D, Pogoda J. Analysis of recurrent events with an associated informative dropout time: Application of the joint frailty model. Statistics in Medicine. 2016;35(13):2195-205.
141. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, et al. A new equation to estimate glomerular filtration rate. Annals of internal medicine. 2009;150(9):604-12.
142. Public Health England. Local action on health inequalities Understanding and reducing ethnic inequalities in health. . 2018.
143. Gov.uk. Population of England and Wales 2020 [Available from: https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/population-of-england-and-wales/latest.
144. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. [PMG9] 5.3 Measuring and valuing health effects. 2013 [Available from: https://www.nice.org.uk/process/pmg9/chapter/the-reference-case#measuring-and-valuing-health-effects.
145. Committee NDJAP. MANAGEMENT OF HEART FAILURE with Reduced Ejection Fraction (HFREF). 2021 April 2021; Review date December 2021.
146. Writing C, Maddox TM, Januzzi JL, Jr., Allen LA, Breathett K, Butler J, et al. 2021 Update to the 2017 ACC Expert Consensus Decision Pathway for Optimization of Heart Failure Treatment: Answers to 10 Pivotal Issues About Heart Failure With Reduced Ejection Fraction: A Report of the American College of Cardiology Solution Set Oversight Committee. J Am Coll Cardiol. 2021;77(6):772-810.
147. DSU N. TSD 18: Methods for population-adjusted indirect comparisons in submissions to NICE. 2016.
148. Singh JSS, Mordi IR, Vickneson K, Fathi A, Donnan PT, Mohan M, et al. Dapagliflozin Versus Placebo on Left Ventricular Remodeling in Patients With Diabetes and Heart Failure: The REFORM Trial. Diabetes Care. 2020;43(6):1356-9.
149. Kato ET, Silverman MG, Mosenzon O, Zelniker TA, Cahn A, Furtado RHM, et al. Effect of Dapagliflozin on Heart Failure and Mortality in Type 2 Diabetes Mellitus. Circulation. 2019;139(22):2528-36.
150. McMurray JJV, Solomon SD, Inzucchi SE, Kober L, Kosiborod MN, Martinez FA, et al. Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction. N Engl J Med. 2019;381(21):1995-2008.
151. Nassif ME, Windsor SL, Tang F, Khariton Y, Husain M, Inzucchi SE, et al. Dapagliflozin Effects on Biomarkers, Symptoms, and Functional Status in Patients With Heart Failure With Reduced Ejection Fraction. Circulation. 2019;140(18):1463-76.
152. Docherty KF, Jhund PS, Inzucchi SE, Køber L, Kosiborod MN, Martinez FA, et al. Effects of dapagliflozin in DAPA-HF according to background heart failure therapy. Eur Heart J. 2020;41(25):2379-92.

153. Komajda M, Anker SD, Cowie MR, Filippatos GS, Mengelle B, Ponikowski P, et al. Physicians' adherence to guideline-recommended medications in heart failure with reduced ejection fraction: data from the QUALIFY global survey. Eur J Heart Fail. 2016;18(5):514-22.
154. Bristow MR, Gilbert EM, Abraham WT, Adams KF, Fowler MB, Hershberger RE, et al. Carvedilol produces dose-related improvements in left ventricular function and survival in subjects with chronic heart failure. MOCHA Investigators. Circulation. 1996;94(11):2807-16.
155. Packer M, Poole-Wilson PA, Armstrong PW, Cleland JG, Horowitz JD, Massie BM, et al. Comparative effects of low and high doses of the angiotensin-converting enzyme inhibitor, lisinopril, on morbidity and mortality in chronic heart failure. ATLAS Study Group. Circulation. 1999;100(23):2312-8.
156. McMurray JJ, Packer M, Desai AS, Gong J, Lefkowitz MP, Rizkala AR, et al. Angiotensin-neprilysin inhibition versus enalapril in heart failure. N Engl J Med. 2014;371(11):993-1004.
157. Swedberg K, Komajda M, Bohm M, Borer JS, Ford I, Dubost-Brama A, et al. Ivabradine and outcomes in chronic heart failure (SHIFT): a randomised placebo-controlled study. Lancet. 2010;376(9744):875-85.
158. Cullington D, Goode KM, Cleland JG, Clark AL. Limited role for ivabradine in the treatment of chronic heart failure. Heart. 2011;97(23):1961-6.
159. Kapelios CJ, Lainscak M, Savarese G, Laroche C, Seferovic P, Ruschitzka F, et al. Sacubitril/valsartan eligibility and outcomes in the ESC-EORP-HFA Heart Failure Long-Term Registry: bridging between European Medicines Agency/Food and Drug Administration label, the PARADIGM-HF trial, ESC guidelines, and real world. Eur J Heart Fail. 2019;21(11):1383-97.
160. Norberg H, Bergdahl E, Lindmark K. Eligibility of sacubitril-valsartan in a real-world heart failure population: a community-based single-centre study. ESC Heart Fail. 2018;5(2):337-43.
161. Ather S, Chan W, Bozkurt B, Aguilar D, Ramasubbu K, Zachariah AA, et al. Impact of noncardiac comorbidities on morbidity and mortality in a predominantly male population with heart failure and preserved versus reduced ejection fraction. J Am Coll Cardiol. 2012;59(11):998-1005.
162. National Institute for Health and Care Excellence. NICE impact cardiovascular disease management 2021 [Available from: https://www.nice.org.uk/about/what-we-do/into-practice/measuring-the-use-of-nice-guidance/impact-of-our-guidance/nice-impact-cardiovascular-disease-management.
163. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, et al. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. New England Journal of Medicine. 2015;373(22):2117-28.
164. Donnan JR, Grandy CA, Chibrikov E, Marra CA, Aubrey-Bassler K, Johnston K, et al. Comparative safety of the sodium glucose co-transporter 2 (SGLT2) inhibitors: a systematic review and meta-analysis. BMJ Open. 2019;9(1):e022577.
165. McEwan P, Darlington O, McMurray JJV, Jhund PS, Docherty KF, Bohm M, et al. Cost-effectiveness of dapagliflozin as a treatment for heart failure with reduced ejection fraction: a multinational health-economic analysis of DAPA-HF. Eur J Heart Fail. 2020;22(11):2147-56.

166. McMurray JJV, Trueman D, Hancock E, Cowie MR, Briggs A, Taylor M, et al. Cost-effectiveness of sacubitril/valsartan in the treatment of heart failure with reduced ejection fraction. Heart. 2018;104(12):1006.
167. Lee D, Wilson K, Akehurst R, Cowie MR, Zannad F, Krum H, et al. Cost-effectiveness of eplerenone in patients with systolic heart failure and mild symptoms. Heart. 2014;100(21):1681-7.
168. National Institute for Health and Care Excellence. Sacubitril valsartan for treating symptomatic chronic heart failure with reduced ejection fraction [TA388] 2016 [Available from: https://www.nice.org.uk/guidance/ta388.
169. National Institute for Health and Care Excellence. Ivabradine for treating chronic heart failure [TA267] 2012 [Available from: https://www.nice.org.uk/guidance/ta267.
170. Scottish Medicines Consortium (SMC). Sacubitril/valsartan (Entresto) 2016 [Available from: https://www.scottishmedicines.org.uk/medicines-advice/sacubitrilvalsartan-entresto-fullsubmission-113216/.
171. Scottish Medicines Consortium (SMC). eplerenone (Inspra) 2012 [Available from: https://www.scottishmedicines.org.uk/medicines-advice/eplerenone-inspra-fullsubmission-79312/.
172. Scottish Medicines Consortium (SMC). ivabradine (Procoralan) 2012 [Available from: https://www.scottishmedicines.org.uk/medicines-advice/ivabradine-procoralan-fullsubmission-80512/.
173. All Wales Medicines Strategy Group (AWMSG). Eplerenone (Inspra) 25mg and 50mg film-coated tablets. Penarth: All Wales Therapeutics and Toxicology Centre (AWTTC), secretariat of the All Wales Medicines Strategy Group (AWMSG). AWMSG Secretariat Assessment Report Advice No. 2712. 2012 [Available from: https://www.crd.york.ac.uk/crdweb/ShowRecord.asp?LinkFrom=OAI&ID=32013000329&LinkFrom=OAI&ID=32013000329.
174. Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. J Am Coll Cardiol. 2000;35(5):1245-55.
175. U.S. Food & Frug Administration. Medical Device Development Tools (MDDT). 2021.
176. Sullivan MD, Levy WC, Russo JE, Crane B, Spertus JA. Summary health status measures in advanced heart failure: relationship to clinical variables and outcome. J Card Fail. 2007;13(7):560-8.
177. Heidenreich PA, Spertus JA, Jones PG, Weintraub WS, Rumsfeld JS, Rathore SS, et al. Health status identifies heart failure outpatients at risk for hospitalization or death. J Am Coll Cardiol. 2006;47(4):752-6.
178. Pokharel Y, Khariton Y, Tang Y, Nassif ME, Chan PS, Arnold SV, et al. Association of Serial Kansas City Cardiomyopathy Questionnaire Assessments With Death and Hospitalization in Patients With Heart Failure With Preserved and Reduced Ejection Fraction: A Secondary Analysis of 2 Randomized Clinical Trials. JAMA Cardiol. 2017;2(12):1315-21.
179. Berg J, Lindgren P, Mejhert M, Edner M, Dahlström U, Kahan T. Determinants of utility based on the EuroQol five-dimensional questionnaire in patients with chronic heart failure and their change over time: results from the Swedish heart failure registry. Value Health. 2015;18(4):439-48.

180. Herbst A, Orso F, Migliorini M, Virciglio S, Tognelli S, Camartini V, et al. ARNI and SGLT2i: a promising association to be used with caution. J Geriatr Cardiol. 2020;17(11):728-32.
181. Ethgen O, Standaert B. Population–versus Cohort–Based Modelling Approaches. PharmacoEconomics. 2012;30(3):171-81.
182. Zaric GS. The impact of ignoring population heterogeneity when Markov models are used in cost-effectiveness analysis. Med Decis Making. 2003;23(5):379-96.
183. Latimer N. NICE DSU Technical Support Document 14: Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data 2013 [Available from: http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf.
184. Sugrue DM, Ward T, Rai S, McEwan P, van Haalen HGM. Economic Modelling of Chronic Kidney Disease: A Systematic Literature Review to Inform Conceptual Model Design. Pharmacoeconomics. 2019;37(12):1451-68.
185. van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health. 2012;15(5):708-15.
186. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35(11):1095-108.
187. Evidera. Statistical Analysis Results for Inputs of the CEM for Empagliflozin in Chronic Heart Failure (Technical Report). Data on File.; 2021.
188. National Institute for Health and Care Excellence (NICE). Position statement on use of the EQ-5D-5L value set for England (updated October 2019). 2019.
189. Sullivan PW, Ghushchyan VH. EQ-5D Scores for Diabetes-Related Comorbidities. Value Health. 2016;19(8):1002-8.
190. Sullivan PW, Ghushchyan V. Preference-Based EQ-5D index scores for chronic conditions in the United States. Med Decis Making. 2006;26(4):410-20.
191. Jesky MD, Dutton M, Dasgupta I, Yadav P, Ng KP, Fenton A, et al. Health-Related Quality of Life Impacts Mortality but Not Progression to End-Stage Renal Disease in Pre-Dialysis Chronic Kidney Disease: A Prospective Observational Study. PLoS One. 2016;11(11):e0165675.
192. Sullivan PW, Slejko JF, Sculpher MJ, Ghushchyan V. Catalogue of EQ-5D scores for the United Kingdom. Med Decis Making. 2011;31(6):800-4.
193. National Health Service. NHS Electronic Drug Tariff. 2021.
194. National health service. National Schedule of NHS costs: 2018/2019 2019 [Available from: https://www.england.nhs.uk/national-cost-collection/.
195. Alva ML, Gray A, Mihaylova B, Leal J, Holman RR. The impact of diabetes-related complications on healthcare costs: new results from the UKPDS (UKPDS 84). Diabet Med. 2015;32(4):459-66.
196. Kerr M, Bray B, Medcalf J, O'Donoghue DJ, Matthews B. Estimating the financial cost of chronic kidney disease to the NHS in England. Nephrol Dial Transplant. 2012;27 Suppl 3(Suppl 3):iii73-80.
197. Kent S, Schlackow I, Lozano-Kühne J, Reith C, Emberson J, Haynes R, et al. What is the impact of chronic kidney disease stage and cardiovascular disease on the annual cost of hospital care in moderate-to-severe kidney disease? BMC Nephrol. 2015;16:65.

198. Eurostat. Harmonised Index of Consumer Prices (HICP): European Commission 2020 [Available from: https://ec.europa.eu/eurostat/web/hicp.
199. Personal Social Services Research Unit. Unit costs of health and social care 2020 2020 [Available from: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2020/.
200. Büyükkaramikli NC, Rutten-van Mölken MPMH, Severens JL, Al M. TECH-VER: A Verification Checklist to Reduce Errors in Models and Improve Their Credibility. PharmacoEconomics. 2019;37(11):1391-408.
201. IQVIA. Review and Validation of Cost-effectiveness Model for JARDIANCE® (Empagliflozin) in Patients with Heart Failure with Reduced Ejection Fraction Version 1.0 [Data on File]. 2021.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Empagliflozin for treating chronic heart failure with reduced ejection fraction [ID3826]

Clarification questions

July 2021

File name	Version	Contains confidential information	Date
ID3826 Empagliflozin clarification questions_FINAL	1.0	Yes	05th August 2021

Notes for company

Highlighting in the template

Square brackets and grey highlighting are used in this template to indicate text that should be replaced with your own text or deleted. These are set up as form fields, so to replace the prompt text in [grey highlighting] with your own text, click anywhere within the highlighted text and type. Your text will overwrite the highlighted section.

To delete grey highlighted text, click anywhere within the text and press DELETE.

Table of Abbreviations

AE	Adverse event
CE	Cost-effective
CEA	Cost-effectiveness analysis
CTR	Clinical trial report
FAD	Final Appraisal Determination
HF	Heart failure
HFrEF	Heart failure with reduced ejection fraction
HHF	Hospitalisation for heart failure
HR	Hazard ratios
ICER	Incremental cost effectiveness ratio
KCCQ-CSS	Kansas City Cardiomyopathy questions – Clinical Summary Score
KCCQ-TSS	Kansas City Cardiomyopathy questions – Total Symptom Score
KCCQ-OSS	Kansas City Cardiomyopathy questions – Overall Summary Score
KM	Kaplan-Meier
MMRM	Mixed model for repeated measures
NA	Not applicable
NYHA	New York Heart Association
OR	Odds ratio
OC-AD	Observed case- after discontinuation of study treatment
OC-OT	observed case on treatment
LOCF	Last observation carried forward
PH	Proportional hazards
RCT	Randomised controlled trial
RR	Rate ratios
RS	Randomised set
SD	Standard deviation
SE	Standard error
SLR	Systematic literature review
SoC	Standard of care
SoC	Standard of care
TS	Treated set
TTD	Time-to-treatment discontinuation
TS	Treated set
QALY	Quality Adjusted Life Year

BI would like to inform NICE that Jardiance received a Marketing Authorisation for HFrEF in Great Britain from the MHRA on 30th July 2021. The Jardiance Summary of Product Characteristics is included in Appendix 4.

The Marketing Authorisation relating to HFrEF is:

> *"Jardiance is indicated in adults for the treatment of symptomatic chronic heart failure with reduced ejection fraction."*

NICE Technical Team clarification question

Priority Question. Please provide the clinical and cost-effectiveness of empagliflozin compared with dapagliflozin. NICE guide to the methods of technology appraisal section 6.2.2 states that when considering the most appropriate comparator the committee will consider a range of factors including established NHS practice in England. dapagliflozin is recommended for use in NHS practice for this same group of patients.

Evidence for empagliflozin vs standard treatment with ACEi, ARBs, BB, MRAs, sacubitril/valsartan is the most relevant for the committee to consider. As discussed in Document B, Section B.1.3.2.3, dapagliflozin does not reflect current standard of care in the UK. The estimated prescribing of dapagliflozin is [redacted] in MQT May 2021 for patients with HF only (1). These data are provided by IQVIA and are derived from THIN, a large and widely used primary care data in the UK. It is updated at the end of each month, and updated estimates can be provided to NICE upon request. We consulted with NHS England during the development of the BIM for the empagliflozin submission, and they were not aware of any alternative data source on the prescribing of dapagliflozin and thus we believe that this is the best available evidence. As future use of treatment and medicines are speculative, we can only reflect the care pathway used today in this submission. This is consistent with NICE guidance and committee discussion TA398 - Section 4.18 - where a specific future scenario was proposed by the manufacturer but rejected by the committee (2).

In addition, clinical experts said that the SGLT2i is another important pillar of care, and that differentiating between empagliflozin and dapagliflozin was not an important consideration.

If dapagliflozin was a relevant comparator, a cost comparison case would be the most appropriate

The ERG and NICE technical team requested a comparison of dapagliflozin vs empagliflozin at the Decision Problem Meeting to prevent any delay in this appraisal. This is because it is included in the Final Scope. Further, at the clarification meeting BI explained that the indirect treatment comparison for empagliflozin vs dapagliflozin showed comparable outcomes (Doc B2.8). Thus, the ERG asked BI to explicitly state in this response why empagliflozin meets the criteria for a cost comparison. A cost comparison case does not require the estimation of an ICER.

Consistent with the criteria for a cost comparison in the Guide to the Processes of Technology Appraisal 2013(3), empagliflozin has demonstrated similar costs and benefits to dapagliflozin, a comparator that has recently received NICE Technology Appraisal Guidance for the same indication (TA679)(4). Justification is described below, and further details are available in Doc B2.8.

- <u>The costs of prescribing empagliflozin are comparable to dapagliflozin</u>
  - The drug acquisition costs of dapagliflozin and empagliflozin are the same. The list price of empagliflozin and dapagliflozin is £36.59 (excluding VAT) and the annual treatment cost is £476.98. The Technology Appraisal Guidance for dapagliflozin is not subject to a Patient Access Scheme (TA679) (4).
  - As empagliflozin and dapagliflozin offer comparable efficacy and have the same dosing frequency and method of administration, it is expected that the resource utilisation is comparable. No additional HCP appointments are required for the initiation of empagliflozin compared to dapagliflozin. Thus, no further analysis is presented. The recommended dose and frequency for empagliflozin and dapagliflozin is 10mg once-daily (5-7).

- <u>The clinical benefits of prescribing empagliflozin are comparable to dapagliflozin</u>
  - A Bucher indirect treatment comparison reported comparable efficacy of empagliflozin versus dapagliflozin, using placebo as the common comparator arm. No statistically significant differences were observed for any outcome tested. For some outcomes, empagliflozin showed a numerical benefit while dapagliflozin showed a numerical benefit for others, but none were statistically significant.
    - The strongest evidence on the similarity of empagliflozin vs dapagliflozin comes from an indirect treatment comparison of the primary composite endpoint. When the DAPA-HF primary composite (HHF, CV-death or urgent HF visits) was compared to the EMPEROR-Reduced primary composite (HHF or CV-death), the HR was not statistically significant ([redacted]).

- Similar trends were observed for key secondary endpoints, including total HHF, renal composite endpoint and KCCQ; however secondary endpoints were not powered to show a statistically significant effect and were not included in hierarchical testing.
- CV-death and all-cause mortality also showed comparable outcomes for empagliflozin vs dapagliflozin.
    - The Bucher ITC estimated the relative efficacy of CV-death to be [redacted]) for empagliflozin vs dapagliflozin; for all-cause mortality the HR was [redacted]
        - Although the ITC reports a slightly lower CV-death and all-cause mortality rate for dapagliflozin than empagliflozin; caution is needed in the interpretation of this result.
        - Neither DAPA-HF or EMPEROR-Reduced were powered to show a statistically significant difference in CV-death or all-cause mortality alone between the intervention and placebo. Further, there were fewer patients and shorter follow-up time in EMPEROR-Reduced than DAPA-HF; and therefore, there is greater uncertainty in where the true point estimate for the HRs lies.
            - This is evident in the widths of the confidence intervals for the HR for CV-death being much larger in EMPEROR-Reduced than DAPA-HF (DAPA-HF, 0.82 (0.69-0.98); EMPEROR-Reduced, 0.92 (0.75-1.12)).
            - A larger sample size reduces the variability in the point estimate. The variability in the CI reduced from [redacted] in EMPEROR-Reduced and DAPA-HF to only 22% when the results of both trials were pooled in a meta-analysis. A pooled meta-analysis estimated a statistically significant HR for CV-death of 0.86 (0.76 – 0.98) for SGLT2i vs SoC. This estimate is likely to be closer to the

true effect size for both empagliflozin and dapagliflozin in HFrEF(8).

- There was a numerical improvement in renal outcomes in patients receiving empagliflozin compared to dapagliflozin, however this was not statistically significant.
    - Patients receiving empagliflozin had [redacted] greater improvement in renal outcomes compared to dapagliflozin [redacted] This was based on a Bucher ITC of empagliflozin vs dapagliflozin using the DAPA-HF definition of worsening renal function. A definition of worsening renal function is provided in **Error! Reference source not found.** (9).

- <u>Contextual considerations for the cost comparison case</u>
    - The evidence presented demonstrates that a cost comparison case for empagliflozin versus dapagliflozin is reasonable. Additionally, the conclusion from the Bucher ITCs that empagliflozin and dapagliflozin offer comparable efficacy across key outcomes for patients with HFrEF is consistent with feedback from UK clinical experts.
    - Given the uncertainty in the estimates from the Bucher ITCs, indicated by the wide confidence intervals, it's unlikely that an estimate of an ICER for empagliflozin vs dapagliflozin in a cost utility framework would support any meaningful decision making. A PSA is only as good as the inputs used, and the inputs from the ITC are highly uncertain. Thus, a PSA scatter plot may reflect underlying random variation rather than being indicative of any real difference between costs and benefits between empagliflozin and dapagliflozin.
    - Having multiple SGLT2is recommended for use by NICE, similar to T2DM, has several benefits:
        - To support continuation of care (i.e., no need to switch T2DM patients already managed with empagliflozin if they develop comorbid HFrEF).
        - Provides more flexibility for local CCGs/budget holders in managing formulary.
        - To overcome potential practical challenges, such as supply chain issues, should they occur.

- Bridge inequalities seen between those with better or worse access to care (Document B, Section B.1.4).
- To allow patient and clinician choice.
- Tailor treatment for patients who are severely renally impaired. Unlike DAPA-HF, EMPEROR-Reduced permitted the inclusion of patients with an eGFR as low as 20mL/min/1.73m^2.
- Management of specific AE's. For example, unlike dapagliflozin, no cases of ketoacidosis were observed in the EMPEROR-Reduced trial (7) (Supplement 2); this assumption is further supported by the EMPA-REG OUTCOME trial in patients with T2DM and established CV disease where very few cases were observed and there was no imbalance between treatment arms (8).

Section A: Clarification on effectiveness data

Baseline characteristics

A1. Priority Question. Please provide the number of patients in each trial arm for EMPEROR-Reduced at baseline who were not receiving the target dose of medical therapy for heart failure (HF) per local guideline.

As shown in Table 1[redacted] of patients in the placebo and empagliflozin trial arms (respectively) were receiving their best tolerated treatment of guideline recommended HF therapies at baseline. This is consistent with NICE Clinical Guideline for HF (NG106) which recommends that patients are optimised on pharmacological therapy, such as ACEi/ARB, BB, MRA, if they continue to be symptomatic (10). No data was collected on the difference between the target dose and the best tolerated dose.

- o

Table 1. Frequency of patients [N(%)] receiving the best tolerated treatment for HF symptoms and other common concomitant diseases or symptoms according to prevailing guidelines (investigator reported).

	Placebo N (%)	Empagliflozin 10mg N (%)	Total N (%)
Number of patients	[redacted]	[redacted]	[redacted]
No	[redacted]	[redacted]	[redacted]
Yes	[redacted]	[redacted]	[redacted]
Missing	[redacted]	[redacted]	[redacted]

Source, CTR: 1245-0110-1611301 Table 4.1 1 (11)
N.B. These are baseline data (trial visit 2) which relates to baseline (week 1) of the study post randomisation.

A2. Priority Question. Please provide baseline KCCQ-CSS, total symptom and overall summary scores in the EMPEROR-Reduced trial by treatment arm for the full trial population.

The KCCQ-CSS, TSS and OSS scores were similar across placebo and empagliflozin at baseline, as shown in Table 2.

Table 2. KCCQ-OSS, TSS, and CSS scores at baseline

	Placebo	Empagliflozin, 10mg
Number of analysed patients	1814	1816
Baseline KCCQ scores, mean (SE)		
KCCQ-CSS	[redacted]	[redacted]
KCCQ-TSS	[redacted]	[redacted]
KCCQ-OSS	[redacted]	[redacted]

Source, CTR (11)Table 15.2.3.6:1 (KCCQ-CSS), RS (OC-AD)[a,b]; Table 15.2.4.26.5:1 (KCCQ-TSS), RS (OC-AD)[a,b]; Table 15.2.4.26.9:1 (KCCQ-OSS), (OC-AD)[a,b];

a. For patients who died, a worse score (score of 0) is imputed at all subsequent scheduled visits after the date of death
b. Based on MMRM adjusted for baseline covariates

A3. Please provide the number of patients and percentages of patients in each quartile for baseline KCCQ-CSS scores by treatment arm in EMPEROR-Reduced for the quartiles:

a) as used in the economic model, and

The table below summarises the number and proportion of patients in each KCCQ-CSS quartile at baseline in the two study arms. Cut-offs to define the KCCQ-CSS quartiles were selected based on the distribution of scores in the combined population at baseline Table 3.

Table 3. Number and percentage of patients in each quartile for baseline KCCQ-CSS as used in the economic model

Baseline KCCQ-CSS	Empagliflozin 10 mg (N = 1,853) *	Placebo (N = 1,852) *	Combined (N = 3,705)
Q1: [0, 55.2)	[redacted]	[redacted]	[redacted]
Q2: [55.2,75)	[redacted]	[redacted]	[redacted]
Q3: [75,89.6)	[redacted]	[redacted]	[redacted]
Q4: [89.6,100]	[redacted]	[redacted]	[redacted]

* Baseline KCCQ-CSS was missing for some patients.

b) for standard quartiles (e.g. KCCQ-CSS score 0-25 etc.).

The ERG clarified that the number and percentage of patients in each standard KCCQ-CSS quartile (KCCQ-CSS 0-25, 26-50, 51-75) was required to enable a comparison with the dapagliflozin NICE appraisal (TA679) (4). However, the quartiles used in the dapagliflozin and empagliflozin cost-effective (CE) models are already similar. In the dapagliflozin CE model submitted to NICE, the quartiles used were 0 to <58, 58 to <77, 77 to <92 and 92 to 100. In the empagliflozin CE model, the quartile cut-offs used were 0 to <55.2, 55.2 to <75, 75 to <89.6 and 89.6 to 100. The ERG clarified that this analysis is not required, but BI should provide a justification for the selection of these cut-off points.

Justification for the selection of cut-off points for KCCQ-CSS

As stated in Doc B3.3.2, "*The model uses a Markov cohort state-transition approach and describes the clinical course of HFrEF using five discrete health*

states defined by quartiles of the baseline distribution of KCCQ-CSS in the combined empagliflozin and placebo treatment arms (KCCQ-CSS quartiles 1 to 4 corresponding to KCCQ-CSS scores of 0 to <55.2, 55.2 to <75, 75 to <89.6, and 89.6 to 100, respectively, with higher score corresponding to a better health status), and death, with health state-specific costs and utilities. The use of quarters vs tertiles was also explored for categorising KCCQ-TSS. Quartiles were found to provide a better fit to the observed data than tertiles while still retaining adequate patient numbers in each subgroup to permit statistically robust analysis and providing sufficient granularity in predicting patient outcomes. Evenly spaced quarters were also rejected (i.e. 0-25, 26-50, 51-75, 76-100) as they did not contain adequate patient numbers in each arm for a robust analysis. Similarly, health states defined by KCCQ-CSS tertiles of the baseline distribution was explored; however the analysis of transition probabilities showed less differentiation between the treatment arms and over time, suggesting loss of sensitivity to differences."

To provide broader context, in the dapagliflozin FAD, the "*ERG noted that cut-offs for the quartiles chosen by the company to measure KCCQ-TSS in the model were arbitrary. But it said it expected that using other cut-offs or approaches to grouping would minimally affect the cost-effectiveness results. The committee concluded that the company's model structure was appropriate for decision making*."(4) Given the similarities in approach, we anticipate similar deliberations for this appraisal.

Outcomes in EMPEROR-Reduced

A4. Priority Question. Please provide the following information regarding change in renal function in EMPEROR-Reduced for each trial arm:

a) **the proportion of patients who have a ≥ 5ml/min reduction in eGFR at each timepoint in Figure 9 in the company submission.**
   Please see response to A4b.
b) **the proportion of patients who have a ≥ 10ml/min reduction in eGFR at each timepoint in Figure 9 in the company submission along.**

Table 4 shows that the proportion of patients who have more than >5 or 10 ml/min/1.73 m^2 decline in eGFR from baseline is comparable for the empagliflozin and placebo arm.

A [redacted] proportion of patient in the empagliflozin arm compared to the SoC had more than a >5 or 10 ml/min/1.73m2 decline in eGFR in week 4 (Table 4). This is due to the mechanism of action of SGLT2 inhibition rather than being indicative of a permanent decline in renal function. A study investigating the role of SGLT2 inhibition with empagliflozin in the kidney showed that by blocking proximal tubule glucose and sodium reabsorption, there is an increase in sodium delivery to the macula densa. This restores tubular-glomerular feedback via afferent arteriolar vasoconstriction. This then reduces renal plasma flow and hyperfiltration which in turn results in a reduction in eGFR due to the reduction in intra-glomerular pressure (a nephroprotective property). This is reflected in the first four weeks of treatment, where patients in the empagliflozin arm had on average a -X[redacted]ml/min/1.73m2 lower eGFR compared to placebo ([redacted]), (CTR, Table 15.2.4.28, RS OC-AD))

Only continued use of empagliflozin offers sustained renal protection. The assumption in the CE model that patients who discontinue empagliflozin receive the same benefits and costs as placebo (Document B, Section B3.6.2, Table 53) is consistent with the mechanism of action described here and was the most conservative approach.

In EMPEROR-Reduced, only [redacted] developed end-stage renal disease in the empagliflozin arm compared to [redacted] patients in the placebo arm [redacted] Further, only [redacted] required chronic dialysis in the empagliflozin arm compared to [redacted] patients in the placebo arm ([redacted] (Request r1090 Table 50.4.1).

Table 4. proportion of patients who have a ≥ 5ml/min and ≥10mL/min reduction in eGFR

Analysis visit	Treatment	Proportion of patients with a ≥5mL/min reduction in eGFR, n (%)	Proportion of patients with a ≥10mL/min reduction in eGFR, n (%)
Number of patients in analysis set	Placebo	[redacted]	
	Empagliflozin (10mg)	[redacted]	
Week 4	Placebo	[redacted]	[redacted]
	Empagliflozin (10mg)	[redacted]	[redacted]
Week 12	Placebo	[redacted]	[redacted]
	Empagliflozin (10mg)	[redacted]	[redacted]
Week 32	Placebo	[redacted]	[redacted]
	Empagliflozin (10mg)	[redacted]	[redacted]
Week 52	Placebo	[redacted]	[redacted]
	Empagliflozin (10mg)	[redacted]	[redacted]
Week 76	Placebo	[redacted]	[redacted]
	Empagliflozin (10mg)	[redacted]	[redacted]

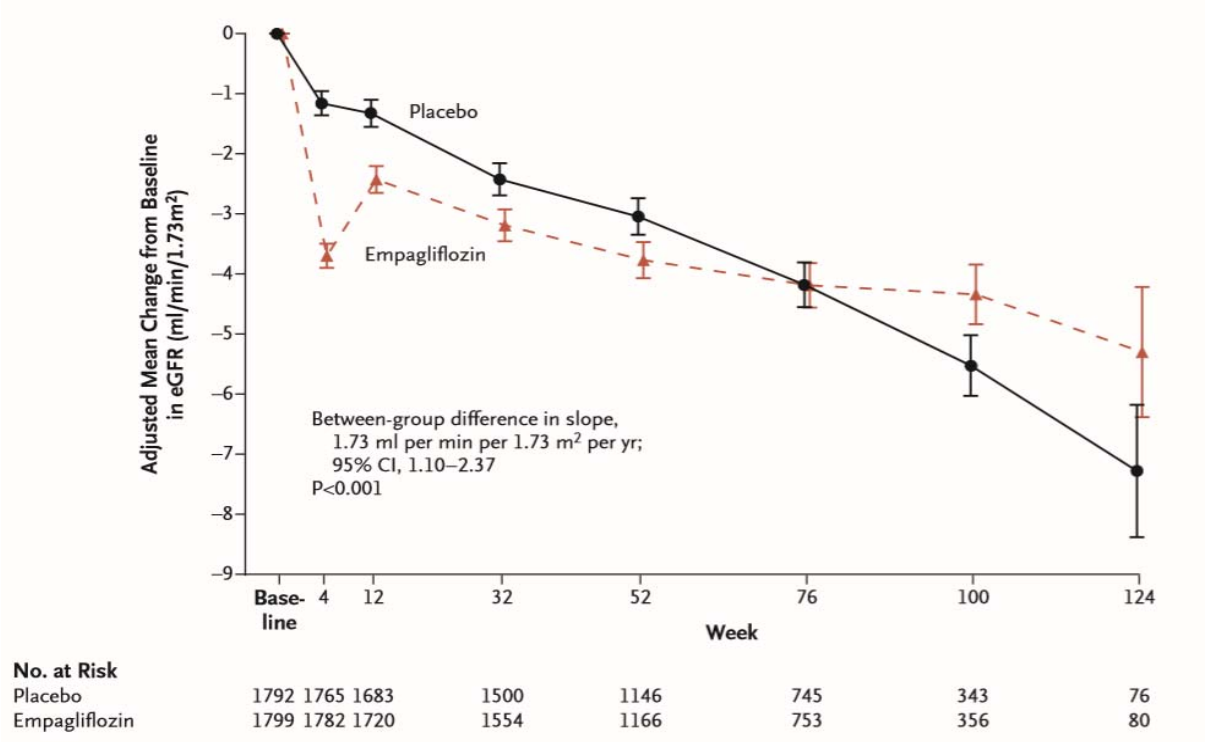
Week 100	Placebo	[redacted]	[redacted]
	Empagliflozin (10mg)	[redacted]	[redacted]
Week 124	Placebo	[redacted]	[redacted]
	Empagliflozin (10mg)	[redacted]	[redacted]
Week 148	Placebo	[redacted]	[redacted]
	Empagliflozin (10mg)	[redacted]	[redacted]

Source: CTR, 1245_121 – Randomised set, observed case – after discontinuation of study medication

c) a table of results to accompany the data presented in Figure 9 of the company submission for mean change from baseline in eGFR.

As stated in Doc B.2.6.2.2, *"the primary analysis included only "on-treatment" data from the treated set (TS) and measurements up to one day after the last intake of study medication. In the empagliflozin arm, the estimated slope was -0.55 ± 0.23 mL/min/1.73m2 per year. In the placebo arm, eGFR declined more steeply over the duration of the treatment period, with an estimated slope of -2.278 ± 0.23 mL/min/1.73m2 per year. The estimated between-arm difference in mean slope was 1.73 mL/min/1.73m2 per year (95% CI, 1.10 – 2.37; p<0.001) (Figure 9). In the randomised set, the adjusted mean eGFR change from baseline to follow-up was 3.3 (95%CI, 1.8-4.8) for empagliflozin versus placebo."* These data are presented in Figure 9 in Doc B and is also shown in Figure 1. As requested, Table 5 provides the data to accompany this figure. Please not that Table 5 and Figure 1 refers to the "on treatment" set whereas Table 4 refers to the observed case after discontinuation of study treatment. This explains why the mean differences from baseline differ slightly.

Figure 1. Changes in the estimated glomerular filtration rate, based on the Treated Set and measurements up to one day after the last intake of study medication



Graph shows the adjusted mean changes from baseline in the eGFR as calculated using the Chronic Kidney Disease Epidemiology Collaboration equation. The bars indicate the standard error. The on-treatment data were analysed with a mixed model for repeated measures (MMRM). Age and baseline eGFR were included as linear covariates, while sex, region, baseline LVEF, baseline diabetes status, last projected visit based on dates of randomisation and trial closure, baseline eGFR according to visit, and visit according to treatment interactions were included as fixed effects. TS, treated set

Table 5. Mean change from baseline in eGFR (mL/min/1.73m^2) over time

Mean change from Baseline to:	Placebo	Empagliflozin (10mg)	Adjusted mean (Empagliflozin vs PBO)
Baseline			-
N	1792	1799	
Mean (SE)	62.3 (0.5)	61.8 (0.5)	
Week 4			
N	1765	1782	-2.5 (0.3) [-3.1 to -2.0] P<0.0001
Mean, (SE), [95%CI]	-1.2 (0.2) [-1.6 to -0.8]	-3.7 (0.2) [-4.1 to -3.3]	
Week 12			
N	1863	1720	-1.1 (0.3) [-1.7 to -0.5], P=0.0005
Mean, (SE), [95%CI]	-1.3 (0.2) [-1.8 to -0.9]	-2.4 (0.2) [-2.9 to -2.0]	
Week 32			
N	1500	1554	-0.8 (0.4) [-1.5 to 0.00], P=0.0407
Mean, (SE), [95%CI]	-2.4(0.3) [-3.0 to -1.9]	-3.2 (0.3) [-3.7 to -2.7]	
Week 52			
N	1146	1166	-0.7 (0.4) [-1.6 to 0.1] P=0.0893
Mean, (SE), [95%CI]	-3.0 (0.3) [-3.6 to -2.5]	-3.8 (0.3) [-4.4 to -3.2]	
Week 76			
N	745	753	0.0 (0.5) [-1.0 to 1.0] P=0.9864
Mean, (SE), [95%CI]	-4.2 (0.4) [-4.9 to -3.5]	-4.2 (0.4) [-4.9 to -3.5]	
Week 100			
N	343	356	1.2 (0.7) [-0.2 to 2.6] P=0.0943
Mean, (SE), [95%CI]	-5.5 (0.5) [-6.5 to -4.5]	-4.3 (0.5) [-5.3 to -3.4]	
Week 124			
N	76	90	2.0 (1.5)

Mean, (SE), [95%CI]	-7.3 (1.1) [-9.5 to -5.1]	-5.3 (1.1) [-7.4 to -3.2]	[-1.1 to 5.0] P=0.1992

Source: CTR, Table 15.2.4.28: 4 (11), adjusted for baseline covariates (MMRM); treated set, observed case – on treatment

d) the proportion of patients who needed acute dialysis.

In the EMPEROR-Reduced trial, patients receiving empagliflozin had a [redacted] reduction in the risk of acute renal injury compared to placebo ([redacted]). The incidence rate for acute renal injury was [redacted] patient years (95% CI [redacted]) for empagliflozin compared to 2.76/100 patient years (95% CI [redacted]) for placebo (CTR, Table 15.2.4.18:1)(11)

The need for dialysis after baseline was reported in EMPEROR-Reduced, however it was not specified whether it was for chronic or acute dialysis. In the empagliflozin arm [redacted] patients ([redacted] required dialysis compared to 20 patients ([redacted]%) in the placebo arm (CTR, Table 15.1.4:12)(11). This was not subject to any significance testing. Dialysis was regarded as chronic if the frequency of dialysis was twice or more per week for at least 90 days.(11)

e) the proportion of patients at any time point who received a renal transplant.

After baseline no patients in either the empagliflozin or the placebo arm of the EMPEROR-Reduced trial required a renal transplant (CTR, Table 15.1.4:12) (11).

A5. Please provide the results of the composite outcome of worsening renal function for each trial arm in EMPEROR-Reduced using the DAPA-HF trial definition, (a sustained decline in the eGFR of 50% or greater, end-stage renal disease, or renal death). For each arm in EMPEROR-Reduced, please provide:

a) the results for the renal composite outcome as defined in DAPA-HF; and
b) the results for each of the individual components of the renal composite outcome as defined in DAPA-HF.

The definition of the renal composite endpoint was more restrictive in EMPEROR-Reduced than DAPA-HF. In both DAPA-HF and EMPEROR-Reduced, the renal composite endpoint included sustained reduction in eGFR as an endpoint. However, unlike DAPA-HF, in EMPEROR-Reduced, measurement of sustained eGFR was contingent on baseline eGFR levels (

Table ***6***). When the renal composite endpoint from EMPEROR-Reduced was re-defined as per DAPA-HF, the clinical outcomes were comparable. Empagliflozin still demonstrated approximately 50% improvement in renal outcomes using either definition (Table 7). This means that the definition of the composite renal outcome does not change the clinical interpretation.

Table 6. Definition of renal composite outcome in EMPEROR-Reduced compared to DAPA-HF

EMPEROR-Reduced (12)	**DAPA-HF (9)**
Time to the first occurrence of any of the components of the composite: • Chronic dialysis. • Renal transplant • Sustained reduction of ≥ 40% eGFR (CKD-EPI)cr or: • Sustained eGFR (CKD-EPI)cr < 15 mL/min/1.73 m² for patients with baseline eGFR ≥ 30 mL/min/1.73 m^2 • Sustained eGFR (CKD-EPI)cr < 10 mL/min/1.73 m² for patients with baseline eGFR < 30 mL/min/1.73 m^2	Time to the first occurrence of any of the components of the composite: • ≥ 50% sustained decline in eGFR or; • reaching end-stage renal disease (ESRD). ESRD was defined as a sustained [≥28 days] eGFR of <15 ml per minute per 1.73 m2, sustained dialysis, or renal transplantation • renal death

Table 7. Renal composite outcome as defined in EMPEROR-Reduced and DAPA-HF

Endpoint	**Placebo n (%) [N=1867]**	**Empagliflozin 10mg n (%) [N=1863]**	**Hazard ratio (95% CI) [Empagliflozin vs dapagliflozin]**
Worsening renal function in EMPEROR-Reduced (as defined in DAPA-HF)			
Renal composite HR	-	-	0.52 (0.29, 0.92):
Individual components[a]			
50% eGFR decline	23 (1.2%)	13/1850 (0.7%)	-
End-stage renal disease	12 (0.6%)	6 (0.3%)	-

Death due to renal cause	2 (0.1%)	2 (0.1%)	-
Composite renal outcomes (as defined in EMPEROR-Reduced) [Randomised set]			
Renal composite HR, 95% CI, P			0.50 (0.32 to 0.77), P=0.0019
Individual components			
Sustained eGFR reduction ≥40% as the first event, N (%)	[redacted]	[redacted]	-
Sustained eGFR <15 mL/min/1.73 m² (baseline ≥30) or <10 mL/min/1.73 m² (baseline <30) as the first event, N (%)	[redacted]	[redacted]	-
Chronic dialysis as the first event, N (%)	[redacted]	[redacted]	-
Renal transplant as the first event, N (%)	[redacted]	[redacted]	-

Source: CTR; Appendix 1 Table 50.4.1[1090], Table 11.1.2.4.3 : 1, Table 15.2.4.1: 1; Table 11.1.2.6 :1; ITC is a data on file

[a]The values provided for the individual components are the total number of patients with the specified event; not the total number of patients with a component as the first event. To estimate the HR for worsening renal function (as defined by DAPA-HF), a cox model was fitted to the dataset ESRD, eGFR (CKD-EPI) sustained reduction ≥50% or death due to renal causes (ITT) using the following covariates: treatment, age, baseline eGFR, sex, baseline LVEF and baseline status.

A6. Priority Question. Please provide a scenario analysis where withdrawals are assumed to not have an event (rather than the imputation method currently used for patients with missing outcome data) for the following outcomes in EMPEROR-Reduced:

a) **total hospitalisation for heart failure (HHF; first and recurrent);**
b) **CV-death;**
c) **All-cause mortality;**
d) **KCCQ-CSS change from baseline;**
e) **treatment discontinuation.**

Section 9.7.1.4 of the clinical trial report (CTR) states "*No data was imputed for safety or for time-to-event endpoints. All efforts were to be made to follow all patients*

until the end of the trial for their survival status and for their survival status and for other endpoints including the primary and key secondary endpoints".

However, imputation was explored as sensitivity analyses for two outcomes:

- The analysis of mean change in KCCQ-CSS score over time was imputed. For patients who died, a worst score (a score of 0) was imputed for all scores planned to be assessed after the date of death.
    - Table 8 reports two analyses of KCCQ-CSS; with and without imputation. Both analyses were MMRM. In both analyses, the empagliflozin arm demonstrated a statistically significant improvement in KCCQ-CSS compared to baseline. This means that assumptions about missing data do not change the clinical interpretation.
    - Further, a cox regression with multiple imputations for patients without primary endpoint events and lost to follow-up before trial completion was undertaken as a scenario analysis in the EMPEROR trial ([redacted] patients in the placebo arm and [redacted] patients in the empagliflozin arm). This is described in Table 16, Section B2.4.1, of the submission under the section *handling drop-outs or missing data*. *Table **9*** show comparable results for the primary composite outcome with and without imputation. Thus, imputation assumptions do not change the conclusion of the trial or the clinical interpretation. It should also be noted that although informative, the primary composite outcome was not used in the cost-effectiveness analysis due to the model design, and thus does not impact on the estimate of cost-effectiveness.

Table 8. Mean change in KCCQ-CSS from baseline to week 12, 32 and 52

Mean change from Baseline to:	**Without imputation**			**With imputation**		
	Placebo	**Empagliflozin (10mg)**	**Diff**	**Placebo**	**Empagliflozin (10mg)**	**Diff**
Baseline						N/A
N	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
Mean (SE)	[redacted]	[redacted]		[redacted]	[redacted]	
Week 12[a]						

Adjusted Mean, SE, [95% CI], P [a]						
Week 32 [a]						
Adjusted Mean, SE, [95% CI], P [a]						
Week 52 [a]						
Adjusted Mean, SE, [95% CI], P [a]						

Source data: Table 15.2.3.6:5, RS, OC-AD without imputation for death; Table 15.2.3.6:1, RS, OC-AD with imputation;
a. adjusted for covariates variables, MMRM; both analyses used the randomised set, observed case after discontinuation

Table 9. Cox regression for time for first hospitalisation HHF or CV-death

	Without imputation (risk set) – base-case analysis			**With multiple imputation (risk set) – scenario analyses**		
	Placebo	**Empagliflozin (10mg)**	**Diff**	**Placebo**	**Empagliflozin (10mg)**	**Diff**
Number of analysed patients, n (100%)						
Number of patients with imputed data						
Number of patients with event, N (%)						
Incidence rate (patients with events/100 patient years), 95% CI						
Comparison with placebo[a], HR [95% CI], P						

Source data: Table 15.2.1.1:1 (without imputation) – Cox model; Table 15.2.1.2:3 (with multiple imputation) – Cox model; a. adjusted for covariates a. randomised set

Abbreviations: NR, not reported

A7. Priority question. Please provide the following for each treatment arm in EMPEROR-Reduced:

a) **the total number of patients with hospitalisations for non-heart failure related reasons (non-heart failure hospitalisation);**
b) **the total number of hospitalisations for non-heart failure related reasons (non-heart failure hospitalisation, first and recurrent);**
c) **a breakdown of the reasons for non-heart failure hospitalisation and the number of patients affected by each reason.**

Non-HHF events are not independent from HHF events and cannot be clearly separated. Clinical experts noted that patients with pre-existing HF that are admitted to hospital may have co-morbidities that could be responsible for their admission. While HF may exacerbate these conditions, it may not be the primary cause for that admission. For example, a patient being admitted for a chest infection may present with similar symptoms to a deterioration in HF - such as shortness of breath. To estimate non-HHF, it is important to consider how HHF was defined in EMPEROR-Reduced. Adjudicated HHF required a committee consensus that the hospitalisation was primarily due to worsening HF (7). Investigator-defined HHF was determined by the trial investigator. Adjudicated time for first HHF occurred in [redacted] patients randomised to placebo and [redacted] patients randomised to empagliflozin ((11), Table 15.2.3.3:1). Investigator-defined time to first HHF occurred in [redacted]patients randomised to placebo and [redacted] patients randomised to empagliflozin ((11), Table 15.2.3.3:2). The total number of adjudicated HHF was [redacted] and [redacted] for placebo and empagliflozin, respectively. The total number of investigator-defined HHF was [redacted] and [redacted] for placebo and empagliflozin, respectively. The analyses of non-HF hospitalisation excluded patients with an investigator-defined HHF.

The proportion of the trial population who experienced a non-HHF event was small, and the proportion experiencing a recurrent non-HHF event was even smaller (Table 10). Further, the proportion of patients in the empagliflozin arm experiencing a

non-HHF was slightly lower than in the placebo arm. The most common reason for non-HF hospitalisation was for cardiac disorders, infections and infestations and these were balanced across arms (Table 11).

As the proportion of non HHF events in the empagliflozin arm was only slightly lower than in the placebo arm, it is unlikely that incorporating non-HHF events into the CE model will have an impact on the ICER, and therefore has not been incorporated. This is a conservative approach. Please refer to Question B10.

Table 10. Proportion of patients with non-HF hospitalisation

	Placebo	Empagliflozin (10mg)
Number of patients in analysis set [a]	[redacted]	[redacted]
Proportion of patients with a non-HF hospitalisation event, N (%)	[redacted]	[redacted]
Proportion of patients with 1 event, N (%)	[redacted]	[redacted]
Proportion of patients with 2 events, N (%)	[redacted]	[redacted]
Proportion of patients with 3 events, N (%)	[redacted]	[redacted]
Proportion of patients with >3 events	[redacted]	[redacted]
Total number of non-HF hospitalisation events	[redacted]	[redacted]

Source: Request r1445, Table 14.1.2.1, Table 14.1.2.2

a. 429 (placebo) and 317 (empagliflozin) had an HHF. Thus the treated set excluding patients with investigator-defined hospitalisation is [redacted]

Table 11. Reasons for non-HF hospitalisation

	Placebo		Empagliflozin (10mg)	
	N, (%)	**Rate/100 patient years**	**N, (%)**	**Rate/100 patient years**
Number of patients in analysis set [a]	[redacted]	[redacted]	[redacted]	[redacted]
Proportion of patients with a non-HHF event, N (%)	[redacted]	[redacted]	[redacted]	[redacted]
Reason for non-HHF				
Cardiac disorder [other than HHF]	[redacted]	[redacted]	[redacted]	[redacted]
Infections and infestations	[redacted]	[redacted]	[redacted]	[redacted]
Nervous system disorders	[redacted]	[redacted]	[redacted]	[redacted]
Gastrointestinal disorders	[redacted]	[redacted]	[redacted]	[redacted]
Injury, poisoning and procedural complications	[redacted]	[redacted]	[redacted]	[redacted]
Respiratory, thoracic and mediastinal disorders	[redacted]	[redacted]	[redacted]	[redacted]
Metabolism and nutrition disorders	[redacted]	[redacted]	[redacted]	[redacted]
Renal and urinary disorders	[redacted]	[redacted]	[redacted]	[redacted]
Vascular disorders	[redacted]	[redacted]	[redacted]	[redacted]
Neoplasms benign, malignant and unspecified (incl cysts and polyps)	[redacted]	[redacted]	[redacted]	[redacted]

Musculoskeletal and connective tissue disorders	[redacted]	[redacted]	[redacted]	[redacted]
Hepatobiliary disorders	[redacted]	[redacted]	[redacted]	[redacted]
General disorders and administration site conditions	[redacted]	[redacted]	[redacted]	[redacted]

Source: Request r1445, Table 14.1.2.3 a. 429 (placebo) and 317 (empagliflozin) had a HHF. Thus the treated set excluding patients with investigator-defined hospitalisation is 1863-429=1434 (PBO); 1863-317=1546

A8. Please provide details of any rationale for the [redacted] [redacted] ([redacted] with empagliflozin versus [redacted]% with placebo).

The proportion of patients who were hospitalised for a non-HF event in EMPEROR-Reduced was comparable ([redacted] (Table 10).

BI were not able to replicate the proportion of patients who experienced a non-HHF estimated by the ERG (18.8% and 13.0%, respectively). We sought clarification from the ERG but did not receive a response before the deadline for submitting these clarification questions.

A9. Please provide results from each trial arm of EMPEROR-Reduced for the outcome KCCQ total symptom score change from baseline at 8 months as used in the indirect treatment comparison reported in Table 23 of the company submission.

The mean KCCQ total symptom score change from baseline at 8 months within both arms of EMPEROR-Reduced is summarised below in Table 12. Empagliflozin had a higher mean change from baseline in KCCQ total symptom score compared with placebo (6.17 compared with 4.53, respectively).

Table 12. Summary of KCCQ total symptom score change from baseline to week 8

Treatment	Number of Patients	KCCQ total symptom score change from baseline to 8 months – mean (sd)
Empagliflozin + SoC	1618	6.17 (19.75)
Placebo + SoC	1569	4.53 (19.75)

Abbreviations: KCCQ, Kansas City Cardiomyopathy Questionnaire; sd, standard deviation; SoC, standard of care

Subgroups

A10. Please provide results including number of events, hazard ratio and 95% confidence intervals for subgroups by age <75 years and ≥ 75 years in EMPEROR-Reduced for the following outcomes:

a) **EMPEROR-Reduced primary composite outcome of time to the first event of adjudicated cardiovascular (CV) death or hospitalisation for heart failure (HHF);**
b) **total hospitalisations for heart failure (first and recurrent);**
c) **CV-death;**
d) **all-cause mortality;**
e) **KCCQ-CSS change from baseline;**
f) **treatment discontinuation.**

Although A10 requests subgroup analyses of EMPEROR-Reduced for the primary and secondary outcomes by age [<75 years and ≥75 years], it was agreed with the ERG during the clarification TC that presenting subgroup analyses by <65; 65–<75; and ≥75 years was acceptable. These subgroup analyses have already been completed by BI in preparation for publication.

The efficacy of empagliflozin in the overall population (ITT) was comparable to the efficacy across subgroups by age. This is observed in similar HRs for the primary composite, the total hospitalisation, and CV and all-cause mortality. A similar trend was observed for KCCQ-CSS where the mean difference across age subgroups was similar to that observed for the overall population. These analyses showed that older patients could expect similar outcomes than younger patients and to the overall population (Table 13). This is an important finding because the average age at diagnosis of HFrEF in the UK is ~71 years (13). The conclusion of these analyses are that, based on age, the efficacy data from EMPEROR-Reduced is generalisable to UK clinical practice.

Treatment discontinuation was not assessed as part of these subgroup analyses. However, since conservative assumptions about treatment discontinuation were

made in the CE model (i.e. those patients who discontinue treatment receive the same benefits as placebo), it is unlikely to have an impact on decision making or the ICER.

Table 13. Primary and secondary outcomes in EMPEROR-Reduced by age subgroup (<65 years, 65-75 years, ≥75 years)

a) EMPEROR-Reduced primary composite outcome of time to the first event of adjudicated cardiovascular (CV) death or hospitalisation for heart failure (HHF)[a]					
Subgroup category	**N with event / N analysed (%)**		**Incidence (event/100 patient years)**		**Hazard ratio (95% CI), P**
	Placebo	**Empagliflozin**	**Placebo**	**Empagliflozin**	
Overall	[redacted]	[redacted]			[redacted]
<65 years	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
65–<75 years	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
≥75 years	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

b) Total HHF (first and recurrent)[b]			
Subgroup category	**Total HHF / N analysed**		**Hazard ratio (95% CI)**
	Placebo	**Empagliflozin**	
Overall	[redacted]	[redacted]	[redacted]
<65 years	[redacted]	[redacted]	[redacted]
65–<75 years	[redacted]	[redacted]	[redacted]
≥75 years	[redacted]	[redacted]	[redacted]

c) CV-death[c]			
Subgroup category	**N with event / N analysed**		**Hazard ratio (95% CI)**
	Placebo	**Empagliflozin**	
Overall	[redacted]	[redacted]	[redacted]
<65 years	[redacted]	[redacted]	[redacted]
65–<75 years	[redacted]	[redacted]	[redacted]
≥75 years	[redacted]	[redacted]	[redacted]

d) All-cause mortality			
Subgroup category	**N with event / N analysed**		**Hazard ratio (95% CI)**
	Placebo	**Empagliflozin**	
Overall	[redacted]	[redacted]	[redacted]
<65 years	[redacted]	[redacted]	[redacted]
65–<75 years	[redacted]	[redacted]	[redacted]
≥75 years	[redacted]	[redacted]	[redacted]

e) KCCQ-CSS change from baseline at week 52[d]			
Subgroup category	**Adjusted mean (SE)**		**Mean difference (95% CI), P**
	Placebo	**Empagliflozin**	
Overall	[redacted]	[redacted]	[redacted]
<65 years	[redacted]	[redacted]	[redacted]
65–<75 years	[redacted]	[redacted]	[redacted]
≥75 years	[redacted]	[redacted]	[redacted]

Source: CTR, Appendix 2, 18.1 [Pg 2-24] (all other outcomes); Appendix 2, Table 37.1.1 (KCCQ).
[a] Cox regression – randomised set, b – estimated as part of a joint frailty model which adjusts for the dependency between the increased risk of death with each subsequent hospitalisation, randomised set; [c] estimated as the cumulative incidence function censoring non-CV-death as a competing risk, randomised set; [d] MMRM, observed case after discontinuation without imputation for death

A11. Please provide subgroup results including hazard ratio and 95% confidence intervals for the Europe geographical region subgroup for patients in each arm of EMPEROR-Reduced for the following outcomes:

a) EMPEROR-Reduced primary composite outcome of time to the first event of adjudicated CV-death or HHF;
b) total HHF (first and recurrent);
c) CV-death;
d) All-cause mortality;
e) KCCQ-CSS change from baseline;
f) treatment discontinuation.

The use of the Europe subgroup to assess generalisability has several limitations, as described below.

- The use of data from the Europe subgroup to assess generalisability is not appropriate because it could contribute to existing ethnic inequalities in health, contrary to the NICE's Social Value Judgments and the Equality Act 2010 (race is one of the protected characteristics) (14). This is why these data were not reported separately in Document B. The Europe subgroup of EMPEROR-Reduced was [redacted] white and therefore not representative of the multi-ethnic UK population, which consists of 86% white, 3.3% black, 7.5% Asian and 3.2% other (15). This difference is even wider in the metropolitan areas of the UK (44.9% white in London) (16). The ITT population of EMPEROR-Reduced, which was 71% white, 6.6% black, 18.1% Asian and 4.2% other (7) is more

generalisable to the ethnically diverse UK population and is, therefore, the population considered in the economic analysis. This is consistent with the committee's perspective in the dapagliflozin appraisal (TA679) (4). The committee recognised that the Europe subgroup was predominantly white, comprised less than half of the overall trial population and may have an absolute risk of complications different to that of patients from the rest of the world (4). The committee concluded that data from the overall DAPA-HF population was acceptable for decision making.

- The use of ITT population for the CE model of empagliflozin in HFrEF is the most statistically robust approach since EMPEROR-Reduced was not powered to evaluate the treatment effect in subgroups. With many subgroup analyses carried out without adjusting the overall significance level of the trial, it is unclear if the results represent spurious findings.

We believe that the ITT population is the most relevant population for the committee to base its recommendation. For the reasons stated above, the Europe subgroup is not generalisable to UK clinical practice and should not be used as the base case in the economic analysis. However, these data have been published. This is why – for transparency – we have reported the results of this subgroup in Table 14 (17). Analyses for all-cause mortality, KCCQ-CSS from baseline and treatment discontinuation has not been conducted.

Table 14. Results of the primary composite outcome, CV-death and total hospitalisation for the Europe subgroup from EMPEROR

Endpoint	N with event / N analysed		Incidence rate/100 patient years		Hazard ratio (95% CI)
	Placebo	Empagliflozin	Placebo	Empagliflozin	
Overall population (17)					
Time to the first event of adjudicated CV-death or HHF	462/1687	361/1863	21.0	15.8	0.75 (0.65 to 0.86), P<0.001
Total HHF (first and recurrent)	553/1867	388/1863	NR	NR	0.70 (0.58 to 0.85), P<0.001
CV-death	202 (10.8)	187 (10.0)	8.1	7.6	0.92 (0.75 to 1.12)

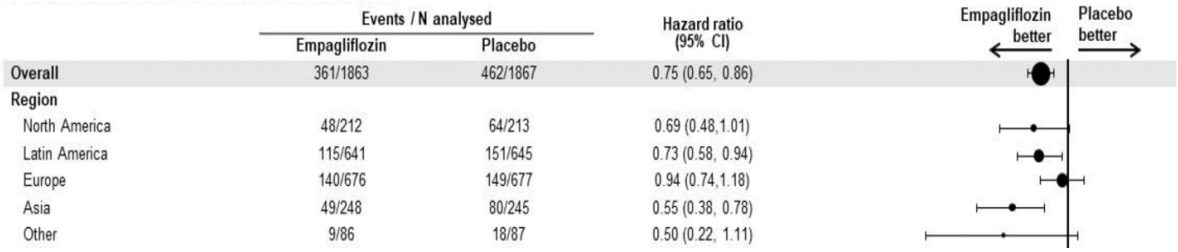
Europe subgroup (17)					
Time to the first event of adjudicated CV-death or HHF	149/677	140/676	16.5	17.5	0.94 (0.74, 1.18)
Total HHF (first and recurrent)[a]	152/677	144/676	15.5	16.3	0.96 (0.70, 1.33)
CV-death	72/677	71/676	7.6	7.7	0.98 (0.71, 1.36)

a) Total HF hospitalisation event rates were derived from an unadjusted negative binomial model
NR, not reported

A12. Please provide subgroup results including number of events, hazard ratio and 95% confidence intervals by geographical region subgroup for patients in each arm of EMPEROR-Reduced for the primary composite endpoint (time to first event of adjudicated CV-death or adjudicated HHF as provided for other subgroups in Figure 16 of the company submission.

As stated in A11, the use of data by geographical region to assess generalisability is not appropriate and could contribute to existing ethnic inequalities in health, contrary to the NICE's Social Value Judgments and the Equality Act 2010 (race is one of the protected characteristics) (14). We report – for transparency - the results of trial by geographical region in **Figure *2*** because these data are publicly available (17). Across all geographical subgroups, empagliflozin demonstrates an improvement in CV-death or HHF compared to the overall population.

Figure 2. Primary composite (CV-death or adjudicated HHF) by region in EMPEROR-Reduced



Treatment discontinuations

A13. Priority Question. Please provide a Bucher independent treatment comparison between empagliflozin and dapagliflozin using the EMPEROR-Reduced and DAPA-HF trials for the outcome of treatment discontinuations.

In EMPEROR-Reduced treatment discontinuation was classified by either a non-fatal adverse event, request by patient or other reasons. In EMPEROR-Reduced a total of [redacted] out of the [redacted] patients assigned to empagliflozin discontinued treatment [redacted] whereas [redacted] patients out of the [redacted] patients assigned to the placebo discontinued treatment [redacted] (11) (Table 10.1.2). In DAPA-HF, treatment discontinuation was also presented as total discontinuations for reasons other than death. In DAPA-HF a total of 249 out of the 2368 patients assigned to, and started, dapagliflozin discontinued treatment (10.5%) whereas 258 patients out of the 2368 patients assigned to, and started, the placebo discontinued treatment (10.9%)(9).

Both empagliflozin and dapagliflozin reduced the odds of treatment discontinuation compared with the placebo, see **Table *15***. The Bucher ITC treatment discontinuation ORs (and 95% CI) for empagliflozin versus dapagliflozin was [redacted] Empagliflozin slightly reduced the odds of treatment discontinuation compared to patients treated with dapagliflozin. It is worth noting that the OR is close to one and the CI contains one suggesting the two treatments have comparable odds of treatment discontinuation.

This Bucher ITC has made several simplifying assumptions

- A subject discontinuing on day 5 is treated the same as a subject discontinuing on day 500.

- Treatment discontinuation excludes death, meaning that patients who die are not counted as discontinuing treatment. This is a simplifying assumption.
    - The estimation of time to treatment discontinuation (TTD) in the CE model is slightly more nuanced. In the original submission, we had censored discontinuations due to fatal AEs at time of discontinuation to avoid double-counting since the model stops treatment at death. Thus, these are not counted as discontinuation in the model, but treatment is effectively stopped for them at death. Since some of patients discontinuing with fatal AE survive past end of treatment, it would be appropriate to capture end of treatment as discontinuations for them. Thus, the updated CE model analyses submitted as part of clarification questions (Appendix 3) was revised to only censor fatal AE discontinuations where deaths occur on the same day. This assumption has been applied to A14, A15, B13, B14, B15 and B16 and is reflected in the updated risk equations (Appendix 1). These updated TTD risk equations do not have a significant impact on the ICER and thus is a minor consideration for the committee. Nonetheless, they were done to ensure accuracy in the evidence presented.

Table 15. Summary of the Bucher ITC result for treatment discontinuation

Endpoint: relative effect measure	**EMPEROR-REDUCED: empagliflozin versus placebo**	**DAPA-HF: dapagliflozin versus placebo**	Bucher ITC: empagliflozin versus dapagliflozin
Treatment discontinuation: OR (95% CI)	[redacted]	[redacted]	[redacted]

Key: CI, confidence interval; OR, odds ratio.

A14. Priority Question. Please provide the annual probability of treatment discontinuation and standard error (SE) for empagliflozin in the EMPEROR-Reduced trial.

A total of [redacted] of the [redacted] patients in the empagliflozin arm discontinued treatment over [redacted] person-years of follow-up. This corresponds to a rate of [redacted] per year (SE=9.8%).

A15. Priority Question. Please provide the following details on-treatment discontinuations for each trial arm in EMPEROR-Reduced:

a) **number of treatment discontinuations excluding death;**
b) **number of treatment discontinuations including death;**
c) **mean with standard deviation (SD) and median time on-treatment excluding death;**
d) **mean (with SD) and median time on-treatment including death.**

The requested information is summarised in Table 16. As stated in A14, [redacted] patients on empagliflozin and [redacted] patients on placebo were discontinued treatment prematurely during the trial. Of these, [redacted] on empagliflozin and [redacted] patients on placebo ([redacted] discontinued due to a fatal adverse event (AE). Not all the fatal AEs occurred at the time of discontinuation. Nearly [redacted] of patients stopping empagliflozin due to a fatal AE died on the same day discontinuation was recorded compared with [redacted] with placebo. The remainder were alive for some time after discontinuation (ranging from [redacted] to [redacted] ays). Duration of treatment is summarised separately for each of these cases.

Table 16. Time on-treatment and reason for discontinuation

Time on-Treatment	Empagliflozin 10 mg			Placebo		
	N	Mean (SD)	Median	N[a]	Mean (SD)	Median
Completed Treatment (All)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Discontinued Treatment (All)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Reason: Fatal AE	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Died at time of discontinuation	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Died after discontinuation	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Reason: Other	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

[a] 4 patients in the placebo arm had missing discontinuation date and are not included in the summary.

PULSE study

A16. Priority Question. Please provide the following baseline characteristics from the PULSE study to supplement those already provided in Appendix O, Table 3 and the comparable baseline characteristics for the placebo and empagliflozin arms of EMPEROR-Reduced:

a) **class of HF medication (e.g. ACE inhibitor, Beta-blocker etc);**
b) **past history of ischaemic heart disease;**
c) **type 2 diabetes;**
d) **New York Heart Association (NYHA) functional class;**
e) **hospitalisation for HF in ≤12 months prior to start of study.**

The additional requested baseline characteristics from PULSE compared to the placebo arm of EMPEROR-Reduced are presented in Table 17. These data show that there are differences in baseline characteristics between PULSE and EMPEROR-Reduced. Compared to PULSE, a higher proportion of patients in EMPEROR-Reduced had a prior HHF in the last 12 months and co-morbid T2DM. In addition, a higher proportion in EMPEROR-Reduced received a beta-blocker, MRA or sacubitril valsartan than the general population.

These results should be considered in context. Prior HHF, co-morbid disease, and treatment are only three factors to consider when drawing conclusions about generalisability of the trial data to UK clinical practice.

There were inherent differences in study design between PULSE and EMPEROR-Reduced, as described below:

- Firstly, there were differences in how HHF was recorded in PULSE and EMPEROR-Reduced, Unlike EMPEROR-Reduced, HHF was not adjudicated by committee. As stated in Question A7, non-HHF is not independent from HHF events and cannot be clearly separated. Patients with pre-existing HF that are admitted to hospital may have co-morbidities that could be responsible for their admission. While HF may exacerbate these conditions, it may not be the primary cause for that admission. For example, a patient being admitted for a chest infection may present with similar symptoms to a deterioration in HF - such as shortness of breath. The implication is that HHF might be recorded as non-HHF, or vice versa, and not accurately captured in PULSE.

- Secondly, not all characteristics in PULSE were recorded to provide a holistic view on generalisability. This means that a holistic view of disease severity could not be obtained. For example, NYHA classification was not well recorded in primary care (Table 17). The observed split between stages 1-4 in the PULSE study should be interpreted with extreme caution, as this is unlikely to be a reliable estimate of the true split across all HFrEF patients included in the study. Further, KCCQ was not recorded in CPRD.

A scenario analysis has been conducted to assess generalisability of the trial data in Question B5 and B6 and considers factors beyond these three baseline characteristics.

Table 17. Baseline characteristics in PULSE compared to the SoC arm of EMPEROR-Reduced

Patients characteristics	PULSE (n=68,780)	EMPEROR SoC (n=1,867) (11)
Hospitalisation for HF in prior 12 months	[redacted]	30.7%
Type 2 diabetes (%)	[redacted]	49.80%
NYHA classification	[redacted]	
Stage 1	[redacted]	0%
Stage 2	[redacted]	75.0%
Stage 3	[redacted]	24.4%
Stage 4	[redacted]	0.6%
Missing	[redacted][a]	0%
Past history of ischemic heart disease (%)	[redacted]	50.7%[b]
ACEi/ARB	[redacted]	68.9%
Beta-Blocker	[redacted]	94.7%
MRA	[redacted][c]	72.6%
Sacubitril Valsartan	[redacted]	20.7%

Source: PULSE as DOF; CTR, Table 10.4.4.1:1
a Not well recorded in primary care. Distribution of actual NYHA amongst those with unknown NYHA in PULSE is not expected to match the distribution amongst those with recorded NYHA stage.
b defined in EMPEROR-Reduced as cause of HF.
c Primary care prescribing data only for PULSE– substantial underestimation expected.

A17. Priority Question. Please provide the results of PULSE including Kaplan-Meier (KM) data with number of patients at risk for the heart failure with reduced ejection fraction (HFrEF) subgroup for the outcomes of:

a) **HHF (first and recurrent);**
b) **CV-death; and**
c) **all-cause mortality.**

A KM plot is only used to show the time to first event rather than first and recurrent. Thus, a KM plot for first and recurrent hospitalisation is not presented. The equivalent nonparametric plot is the mean cumulative function, shown in Figure 3. This represents the average number of HHF experienced since index. The requested KM plots for all-cause and CV-death are shown in Figure 5 and Figure 7, respectively.

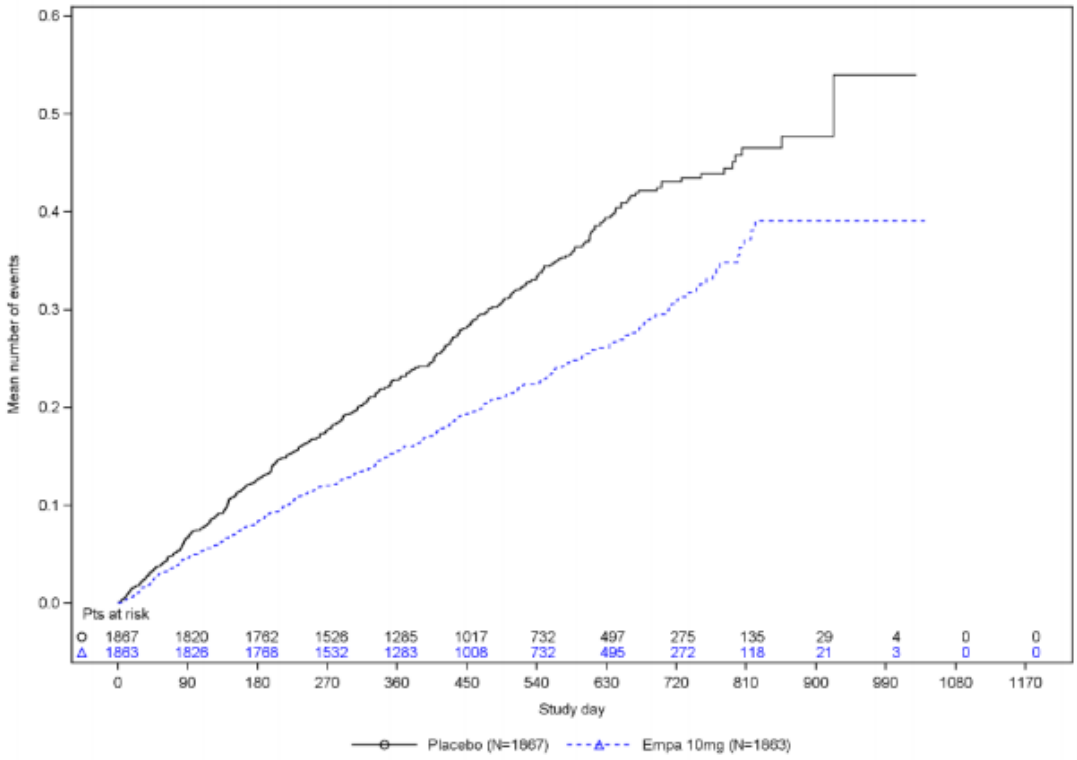
The cumulative number of HHF events (first and recurrent) was lower for PULSE than EMPEROR-Reduced (Figure 3 and Figure 4). In PULSE, the age and sex adjusted HHF rate was 10.2/100 patient years (13) compared to 22.3/100 patient years in the placebo arm of EMPEORER-Reduced (Appendix 1). These rates are for first and recurrent HHF. Please refer to Question B8 for further discussion about the assumptions made about HHF rates in the CE model.

The probability of dying of either HHF or another cause was comparable in EMPEROR-Reduced and PULSE (Figure 5, Figure 6, Figure 7, Figure 8). In PULSE(13), the age adjusted CV-death rate for HFrEF was 6.1/100 patient years. In comparison, in EMPEROR-Reduced(7), the CV death rate in the placebo arm was 8.1/100 patient years. For all-cause mortality, the rate in the PULSE study was 12.7/100 patient year (age and sex adjusted)(13) compared to a rate of 10.7/100 patient years in EMPEROR-Reduced was (7).

The conclusion from these comparisons is that although there might be differences in the baseline characteristics between the trial population and the UK clinical practice, the all cause-death outcomes observed in PULSE and EMPEROR-Reduced were broadly comparable, indicating that the difference in HHF rates might be due to how they were recorded.

Figure 3. Mean cumulative function for first and recurrent HHF in HFrEF patients in PULSE

Figure 4. Mean cumulative incidence of adjudicated HHF (first and recurrent) in EMPEROR-Reduced

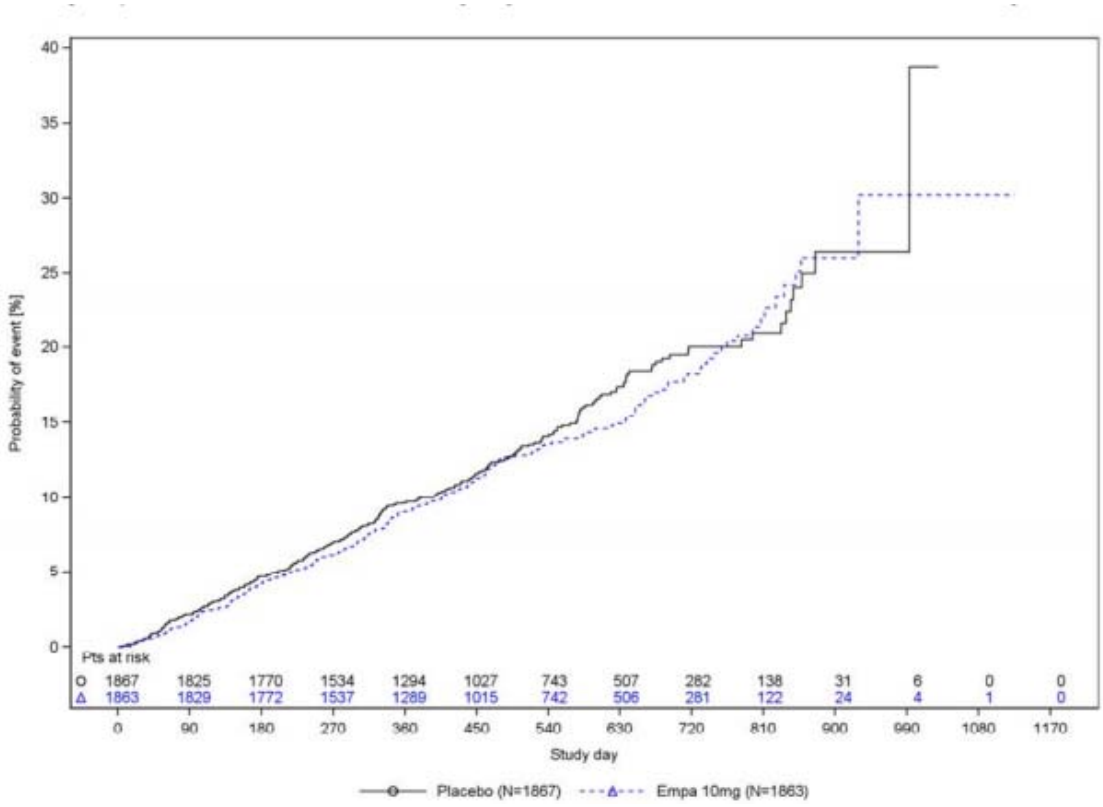


Source: CTR, Figure 11.1.2.1.1:1

Figure 5. All-cause mortality KM for HFrEF patients in PULSE

[redacted]

Figure 6. Kaplan-Meier time to all-cause mortality in EMPEROR-Reduced

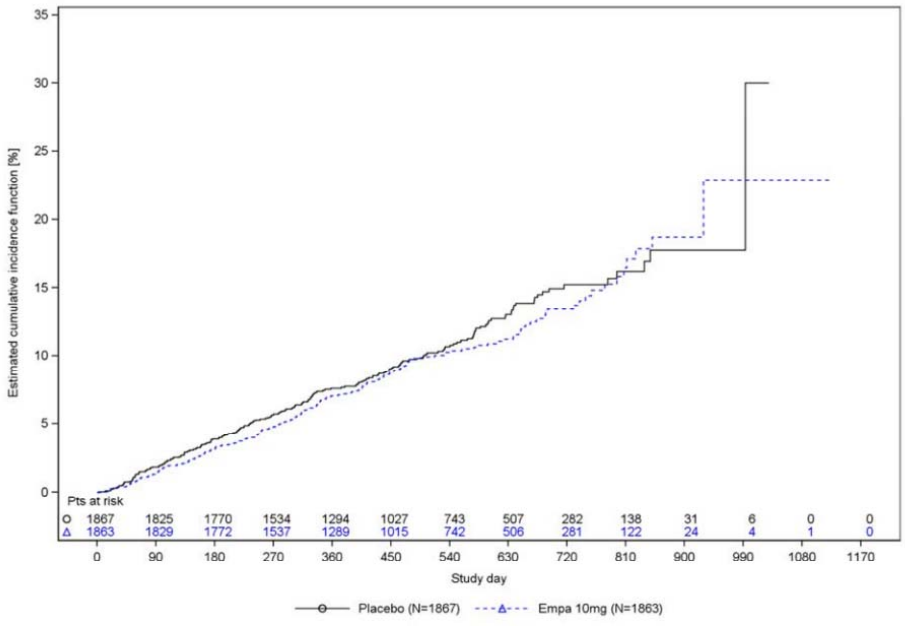


Source: CTR, Figure 15.2.3.2:1; Randomised set

Figure 7. Cardiovascular mortality KM for HFrEF patients in PULSE

[redacted]

Figure 8. Time to adjudicated CV-death, estimated cumulative incidence function (considering non-CV death as a competing risk) in EMPEROR-Reduced



Source: CTR, Figure 11.1.2.4.2:1

Systematic literature review

A18. It is noted that 45 studies met the inclusion criteria for the systematic literature review although data from only 2 studies are presented in the company submission. Please provide a table to summarise the rationale for not presenting data from each of the remaining 43 included studies in the systematic literature review.

A systematic literature review (SLR) was conducted to identify from the published literature randomised controlled trial (RCT) evidence that met the broad PICOS criteria for studies of empagliflozin in HFrEF. The SLR identified a total of 45 studies, extracted from 356 publications, and this included three studies that reported outcomes with empagliflozin (7, 8, 18) and four studies describing efficacy outcomes with dapagliflozin (9, 19-21). The initial SLR was broad so it could be used across multiple countries and HTA submissions (After the Final Scope was finalized for this submission, the 45 included studies were assessed against the decision problem. Most were excluded because the intervention did not include empagliflozin or the outcome was not relevant. The reasons for excluding 43 of these studies are presented below.

Table *18*).

After the Final Scope was finalized for this submission, the 45 included studies were assessed against the decision problem. Most were excluded because the intervention did not include empagliflozin or the outcome was not relevant. The reasons for excluding 43 of these studies are presented below.

Table 18. Reason for exclusion from the clinical SLR

No.	Study name	Trial name	Reason for exclusion
1.	Armstrong 2020 (22)	VICTORIA (Vericiguat Global Study in Patients with HFrEF)	Intervention
2.	Carbone 2020(23)	CANA-HF study	Intervention
3.	Cosentino 2020(24)	VERTIS-CV (eValuation of ERTugliflozin effIcacy and Safety CardioVascular outcomes trial)	Intervention
4.	Edelmann 2020(25)	OUTSTEP-HF (randOmised stUdy using accelerometeTry to compare Sacubitril/valsarTan and Enalapril in Patients with Heart Failure)	Intervention
5.	Gao 2020(26)	NR	Intervention

6.	Jensen 2020(27)	EMPIRE-HF	Outcome
7.	Khandwalla 2019(28)	AWAKE-HF	Intervention
8.	Mordi 2020(29)	RECEDE-CHF	Outcome
9.	Mullasari 2020(30)	PROFICIENT (PROlonged Release Formulation of Ivabradine OnCe-DaIly in HEart Rate ManagemeNT)	Intervention
10	Panagov 2020(31)	NR	Intervention
11.	Singh 2020(19)	REFORM (Research into the Effect Of SGLT2 inhibition on left-ventricular Remodelling in patients with HF and diabetes Mellitus)	Outcome
12.	Tanaka 2020(32)	CANDLE	Intervention
13	Desai 2019(33)	EVALUATE-HF	Intervention
14	Felker 2020(34)	STANDUP-Imaging	Intervention
15	Kang 2019(35)	PRIME study (Pharmacological Reduction of Functional, Ischaemic Mitral Regurgitation)	Intervention
16	Kato 2019(20)	DECLARE-TIMI 58 (dapagliflozin Effect on Cardiovascular Events in Type 2 Diabetes Mellitus 58)	Population
17.	Nassif 2019(36)	DEFINE-HF (dapagliflozin Effects on Biomarkers, Symptoms and Functional Status in Patients with HF with Reduced Ejection Fraction)	Outcome
18.	Tsutsui 2019(37)	J-SHIFT Study	Intervention
19.	Villacorta 2019(38)	NR	Intervention
20.	Raja 2018(39)	NR	Intervention
21.	NCT02788656 2016(40)	PARENT (Pulmonary Artery Pressure Reduction With ENTresto)	Intervention
22.	Sallam 2016(41)	NR	Intervention
23.	Senni 2016(42)	TITRATION study	Intervention
24.	Teerlink 2016(43)	COSMIC-HF (Chronic Oral Study of Myosin Activation to Increase Contractility in Heart Failure)	Intervention
25.	Tsutsui 2016(44)	NR	Intervention
26.	Abdel-Salam 2015(45)	NR	Intervention
27.	Amosova 2015(46)	NR	Intervention
28.	Gheorghiade 2015(47)	SOCRATES-REDUCED (Soluble Guanylate Cyclase Stimulator in HFrEF Study)	Intervention
29.	Greenberg 2015(48)	NR	Intervention
30.	Lopatin 2015(49)	NR	Intervention
31.	Ordu 2015(50)	NR	Intervention
32.	Chaudhari 2014(51)	NR	Intervention
33.	McMurray 2014(52)	PARADIGM-HF (Prospective Comparison of ARNI with ACEI to Determine Impact on Global Mortality and Morbidity in Heart Failure)	Intervention
34.	Moiseev 2011(53)	NR	Intervention
35.	Volterrani 2011(54)	CARVIVA HF	Intervention
36.	Sarullo 2010(55)	NR	Intervention
37.	Swedberg 2010(56)	SHIFT (Systolic HF Treatment with the If Inhibitor Ivabradine Trial)	Intervention
38.	Fox 2008(57)	BEAUTIFUL (morBidity-mortality EvAlUaTion of the If inhibitor ivabradine in patients with coronary disease and left-ventricULar dysfunction)	Intervention

39.	Taylor 2004(58)	A-HeFT (African-American HF Trial)	Intervention
40.	Cohn 1991(59)	V-HeFT II (Vasodilator-HF Trial II)	Intervention
41.	Lin 1991(60)	NR	Intervention
42.	Schofield 1991(61)	NR	Intervention
43.	Cohn 1986(62)	V-HeFT I (Vasodilator-HF Trial I)	Intervention

A19. Priority Question. It is reported in the company submission that three studies of empagliflozin and four studies of dapagliflozin were included in the systematic literature review although only one study for each drug was included in the independent treatment comparison. Please provide a detailed rationale for not including the remaining two studies of empagliflozin and three studies of dapagliflozin.

The SLR identified three empagliflozin studies and four dapagliflozin studies. Only EMPEROR-Reduced and DAPA-HF were included in the ITC because they included the outcomes relevant to the economic model and were the highest quality of evidence. Section B2 of Document B states that the manufacturer should only include studies that are used to inform the economic model.

Table 19 summarises the reasons for using the EMPEROR-Reduced and DAPA-HF studies as the evidence base for ITCs of empagliflozin versus dapagliflozin, respectively and why the remaining studies were excluded.

Table 19: Summary of study inclusion/exclusion in the evidence base

Study	Included/Excluded	Primary reasons for exclusion
EMPEROR-REDUCED	Included	NA
EMPIRE-HF	Excluded	Excluded based on reduced study duration and study objective
RECEDE-CHF	Excluded	Excluded based on reduced study duration and irrelevant outcomes for the CE model
DAPA-HF	Included	NA
REFORM	Excluded	Excluded based on small patient numbers and different focus of analysis and irrelevant outcomes for the CE model
DECLARE-TIMI 58	Excluded	Excluded based lack of randomisation (within the subgroup) and heterogeneity due to all patients being T2DM
DEFINE-HF	Excluded	Excluded based on small study duration and study objective

Key: NA, not applicable.

EMPIRE-HF and RECEDE-HF were excluded because the outcomes did not adequately inform the key outcomes in the CE model, the follow up was short and the sample size was small compared to EMPEROR-Reduced (Table 19, Table 20). There were additional reasons for their exclusion .All studies had the placebo as the comparator arm however the intervention arm slightly differed, RECEDE-CHF used a 25mg dose of empagliflozin (this is not the licensed dose for HFrEF) whereas the remaining two trials were a 10mg dose of empagliflozin. Each study was in a different phase with EMPIRE-HF being Phase II, EMPEROR-Reduced Phase III and RECEDE-CHF Phase IV. Multi-centre studies provide a more robust sample than single centre studies. All studies were double blinded however EMPEROR-Reduced, and EMPIRE-HF were multi-centre studies whereas RECEDE-CHF was a single-centre study.

Table 20: Trial design for the empagliflozin studies

	EMPEROR-REDUCED	EMPIRE-HF	RECEDE-CHF
Intervention (N)	Empagliflozin (10 mg qd) (N = 1863)	Empagliflozin (10 mg qd) (N = 95)	Empagliflozin (25 mg qd) (N = 23)
Comparator (N)	Placebo (N = 1867)	Placebo (N = 95)	Placebo (N = 23)
Objective	To evaluate the effects of empagliflozin on the morbidity and mortality of patients with established HFrEF, with or without type 2 diabetes	To investigate the effect of empagliflozin on NT-proBNP in patients with HFrEF	To assess the diuretic and natriuretic effect of empagliflozin in combination with loop diuretics in patients with type 2 diabetes and chronic HF
Phase	III	II	IV
Method of blinding	Double-blind	Double-blind	Double-blind
Study centres	Multi-centre (Europe, North America, Latin America, Asia, Other)	Multi-centre (Denmark)	Single-centre (Scotland)
Primary outcome	Time to CV-death or adjudicated hospitalisation	Between-arm difference in the change of plasma concentrations of NT-proBNP [Time Frame: 90 days]	change in 24-hour urinary volume from baseline to week 6
Duration of study[a]	Median follow-up: 69.5 weeks (16 months)	Study Duration: 12 weeks	Study Duration: 15 weeks

Key: HF, heart failure, HFrEF, heart failure with a reduced ejection fraction; N, number of patients; NT-proBNP, N-terminal pro-b-type natriuretic peptide; qd, once a day; SoC, standard of care.
Note: a – Weeks calculated using a google calculator if not provided

DECLARE-TIMI, REFORM, and DEFINE-HF were excluded because the population did not fully align with the Final scope, the follow up was short or the sample size was small (Table 21). There were additional reasons for their exclusion. All studies were double blinded however REFORM was a single-centre study and the remaining three studies were multi-centre. Another difference between the studies is in their duration. DAPA-HF had a median follow-up duration of 18.2 months, in comparison, REFORM was shorter at 12 months and DEFINE-HF was shorter again, with study duration of 15 weeks. On the other hand, DECLARE-TIMI 58 was longer with median follow-up duration of 4.2 years and study duration of 6 years. The DEFINE-HF study was excluded from further consideration due to the short follow-up being inappropriate for the outcomes of interest. The REFORM study was excluded from the ITCs due to the small sample size and different focus in terms of outcomes. The DECLARE-TIMI study was excluded from the ITC as the HFrEF subgroup could not be considered a randomised comparison (HFrEF was not a stratification factor) and all patients had T2DM

Table 21: Trial design for the dapagliflozin studies

	DAPA-HF	REFORM	DECLARE-TIMI 58	DEFINE-HF
Intervention (N)	dapagliflozin (10 mg qd) (N = 2373)	dapagliflozin (10 mg qd) (N = 28)	dapagliflozin (10 mg qd) (N = 318)[b]	dapagliflozin (10 mg qd) (N = 131)
Comparator (N)	Placebo (N = 2371)	Placebo (N=28)	Placebo (N = 353)[b]	Placebo (N = 132)
Population	HFrEF	T2DM and HF	T2DM	HFrEF
Objective	To prospectively evaluate the efficacy and safety of dapagliflozin in patients with HFrEF, regardless of the presence or absence of diabetes	To determine the cardiac effects of dapagliflozin in patients with HF and type 2 diabetes mellitus on LV remodelling using cardiac MRI and to help explain the substantial improvements in HF outcomes seen in large clinical trials	To examine the efficacy and safety of dapagliflozin according to baseline HF status and systolic LVEF	To evaluate the effect of dapagliflozin on biomarkers, symptoms, and functional status in patients with HFrEF
Phase	III	IV	III	IV
Method of blinding	Double-blind	Double-blind	Double-blind	Double-blind
Study centres	Multi-centre (Europe, North America, Latin America, Asia Pacific)	Single-centre (Scotland)	Multi-centre (Europe, North America, Latin America, Asia Pacific)	Multi-centre (US)
Primary outcome	Composite of worsening HF (hospitalisation or an urgent visit resulting in intravenous therapy for HF) or cardiovascular death	Change in left-ventricular end-systolic volume End-systolic volume (LVESV)	The primary safety outcome was MACE (defined as cardiovascular death, myocardial infarction, or ischaemic stroke)	Dual primary end points were (1) the average of 6- and 12-week mean NT-proBNP and (2) a composite of the proportion of patients that achieved a meaningful improvement in health status (≥5-point increase in average of 6- and 12-week KCCQ-OS) or NT-proBNP (≥20% decrease in average of 6- and 12-week NT-proBNP)

Duration of study[a]	Median follow-up: 79.08 weeks (18.2 months)	Study Duration: 52.14 weeks (1 year)	Study Duration: 312.86 weeks (6 years) Median follow-up: 219 weeks (4.2 years)	Study Duration: 15 weeks

Key: HF, heart failure, HFrEF, heart failure with a reduced ejection fraction; LV, left-ventricular; LVEF, left-ventricular ejection fraction; MRI, magnetic resonance imaging; N, number of patients; qd, once a day; SoC, standard of care.
Note: a – Weeks calculated using a google calculator if not provided
b – HFrEF subgroup of trial population

A20. Please provide a detailed quality assessment for the DAPA-HF trial, (the study used in the independent treatment comparison) similar to the quality assessment provided in Table 17 of the company submission for EMPEROR-Reduced.

A summary of the quality assessment of EMPEROR-Reduced and DAPA-HF trial is shown in Table 22.

Table 22. Results of quality assessment of trials used in independent treatment comparison

	EMPEROR-Reduced (NCT03057977)	DAPA-HF (NCT03036124)
Was randomisation carried out appropriately?	Yes. Randomisation was performed by using a permuted block design with a computer pseudo-random number generator.	Yes. Randomisation was performed in accordance with the sequestered, fixed-randomisation schedule, with the use of balanced blocks to ensure an approximate 1:1 ratio of the two regimens.
Was the concealment of treatment allocation adequate?	Yes. An Interactive Response Technology System (voice response or web response) was used to determine treatment assignment.	Yes. An Interactive voice response or web response system was used to determine treatment assignment.
Were the arms similar at the outset of the study in terms of prognostic factors?	Yes. Demographic and patient characteristics were well balanced between the two treatment arms at baseline, and randomisation was stratified by geographical region, diabetes status and eGFR at screening.	Yes. Demographic and patient characteristics were well balanced between the two treatment arms at baseline, and randomisation was stratified by Randomisation was stratified based on a diagnosis of type 2 diabetes (i.e., an established diagnosis or a glycated haemoglobin level of ≥6.5% [≥48 mmol per mole]) confirmed at screening.
Were the care providers, participants and outcome assessors blind to	Yes. This was a double-blind study. An Endpoint Adjudication Committee evaluated all reported and potential clinical events in a	Yes. This was a double-blind study. The dapagliflozin tablets and the respective placebo tablets were identical in size, colour, smell, and taste. The bottles with investigational product were

	EMPEROR-Reduced (NCT03057977)	DAPA-HF (NCT03036124)
the treatment allocation?	manner blinded to the treatment assignment.	labelled with unique identification numbers. No member of the extended AstraZeneca study team, personnel at study sites, or any CRO handling study data will have access to the randomisation scheme during the study. The sponsor (AstraZeneca) personnel or delegate generating the randomisation scheme and the supply chain study management may be able to access the randomisation scheme as appropriate.
Were there any unexpected imbalances in drop-outs between arms? If so, were they explained or adjusted for?	No. Proportion of patients who discontinued study treatment was low and well balanced between the two treatment arms.	No. There were no unexpected drop-outs between arms. Proportion of patients who discontinued were balanced between the two treatment arms.
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No. All outcomes specified in the study protocol were reported in the clinical study report.	No. All outcomes specified in the study protocol were reported in the study publication.
Did the analysis include an intention to treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes. Efficacy analysis were performed in the randomised set.	Yes. Efficacy analysis were performed in the randomised set.

Section B: Clarification on cost-effectiveness data

For any scenarios requested in Section B, please ensure these are implemented as user selectable options in the economic model so that these can be combined. Furthermore, if the company chooses to update its base-case results, please ensure that cost-effectiveness results, sensitivity and scenario analyses incorporating the revised base-case assumptions are provided with the response along with a log of changes made to the company base-case.

New question added by the ERG:

From page 100 in the clinical study report, it seems that approximately 68% of placebo patients and 61% of empagliflozin patients were still on-treatment/being observed at 52 weeks in the EMPEROR trial. The ERG also notes that the mean time on-treatment in trial was 63 weeks, and mean follow-up time was 67 weeks. Please provide:

a) Number of observations at week 12, week 32, week 52, end of treatment visit, and follow-up visit in the KCCQ-CSS dataset used in the economic analysis to estimate the transition probabilities between the KCCQ-CSS states of the model (without imputed values);

The number of observations with available KCCQ-CSS scores used in transition probability analyses are provided in Table 23.

b) **The mean and respective standard deviation (SD) KCCQ-CSS data (without imputed values) at week 12, week 32, week 52, end of treatment visit, and follow-up visit in the KCCQ-CSS dataset used in the economic analysis to estimate the transition probabilities between the KCCQ-CSS states of the model;**

KCCQ-CSS transition matrices were derived from observations taken while patients were still alive and followed in the study, as shown in Table 23.

Table 23. Mean and SD for KCCQ-CSS data (without imputed values) as used in the economic model.

	Placebo		Empagliflozin 10 mg	
KCCQ-CSS Score without Imputation	**N**	**Mean (SD)**	**N**	**Mean (SD)**
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Week 12	[redacted]	[redacted]	[redacted]	[redacted]
Week 32	[redacted]	[redacted]	[redacted]	[redacted]
Week 52	[redacted]	[redacted]	[redacted]	[redacted]

c) The number of observations and mean (and respective SD) KCCQ-CSS data (with imputed values) for week 12, week 32 and week 52 underpinning the KCCQ data used in the analysis to estimate the transition probabilities in the model;

The requested information is summarised in Table 24.

Table 24. Mean and SD for KCCQ-CSS data (with imputed values) as used in the economic model.

	Placebo		Empagliflozin 10 mg	
KCCQ-CSS Score with Imputation	**N**	**Mean (SD)**	**N**	**Mean (SD)**
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Week 12	[redacted]	[redacted]	[redacted]	[redacted]
Week 32	[redacted]	[redacted]	[redacted]	[redacted]
Week 52	[redacted]	[redacted]	[redacted]	[redacted]

* The higher number of observations at week 12 are due to records from patients with missing scores at baseline. These patients contribute data on transitions from week 12 onwards and were kept in the analyses.

d) The details, data used and the results of the exploratory analysis conducted by the company, described in the Evidera appendix (page 6) which concluded that "KCCQ-CSS health states tend to change early on after start of treatment and stabilise fairly early

The assessment that changes in KCCQ-CSS tend to occur early was based on data presented in Table 24, Table 25 and

Table ***26*** (Appendix 2, Table 2,3,4). The mean KCCQ-CSS scores over time are relatively constant at weeks 12, 32 and 52 (Table 24). This is consistent with and without imputation, as shown in response b and c. This is also illustrated in Figure 9 showing the change in KCCQ-CSS scores over time (without imputation). Similarly, the monthly transition probabilities (Table 25,

Table ***26***) show that that probability of remaining in the same quartile is higher for Week 12-32 and Week 32-52 transitions ([redacted] across both treatment arms) compared with Baseline-Week 12 ([redacted] across both treatment arms).

Figure 9. Change in KCCQ over time from EMPEROR-Reduced

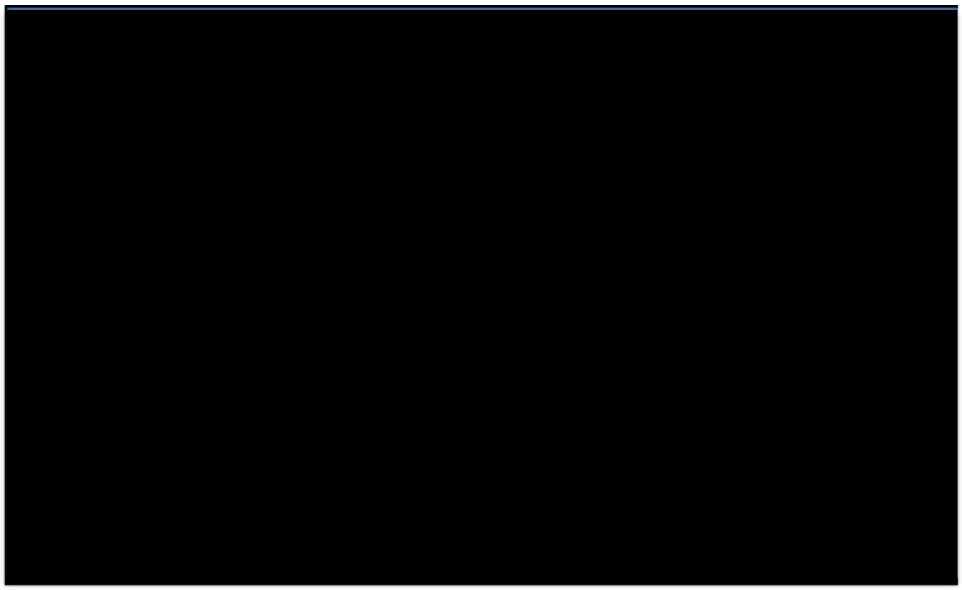


Table 25. Monthly Transition Probability Matrices for KCCQ-CSS Health States in the Empagliflozin 10 mg Arm

From/To	[0, 55.2)	[55.2,75)	[75,89.6)	[89.6,100]	Percent Declined	Percent Improved
Baseline to Week 12						
[0, 55.2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[55.2,75)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[75,89.6)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[89.6,100]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 12 to Week 32						
[0, 55.2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[55.2,75)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[75,89.6)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[89.6,100]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 32 to Week 52						
[0, 55.2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[55.2,75)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[75,89.6)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[89.6,100]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

* Expressed as the probability of moving from the current KCCQ-CSS level to each of the four possible levels over the next month. Probabilities add to 1 in each row. The numbers in parentheses reflect counts of patients making those transitions over the full period represented in each section (baseline-week 12, 12-32, 32-52).

Table 26. Monthly Transition Probability Matrices for KCCQ-CSS Health States in the Placebo Arm

From/To	[0, 55.2)	[55.2,75)	[75,89.6)	[89.6,100]	Percent Declined, %	Percent Improved, %
Baseline to Week 12						
[0, 55.2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[55.2,75)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[75,89.6)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[89.6,100]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 12 to Week 32						
[0, 55.2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[55.2,75)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[75,89.6)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[89.6,100]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 32 to Week 52						
[0, 55.2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[55.2,75)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[75,89.6)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[89.6,100]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

* Expressed as the probability of moving from the current KCCQ-CSS level to each of the four possible levels over the next month. Probabilities add to 1 in each row. The numbers in parentheses reflect counts of patients making those transitions over the full period represented in each section (baseline-week 12, 12-32, 32-52).

Mortality

B1. Priority question. Please include a tab in the Excel model where all the alternative survival distributions provided in the economic model to estimate CV-death (using the joint modelling approach) are plotted against the respective KM data on CV-death from EMPEROR-Reduced. Please ensure the KM data and fitted survival data (underpinning the plots) are provided in the same tab.

Two new tabs (i.e., KM Curves - CVM, KM Curves - ACM) were added into the CE model for CV-death and all-cause mortality, respectively, and plot alternative survival distribution against KM data over time. The distributions depicted in each tab are linked to the value the user selects from the dropdown in F90 in the "Clinical Inputs" tab and update automatically whenever a new fit is chosen. Each plot incorporates KM data and parametric fits (the solid and dashed series, respectively) for both empagliflozin with SoC and SoC alone (the blue and yellow series, respectively).

B2. Priority question. Please provide the p-values associated with [redacted] Please provide the same p-values for all the different distributions considered (exponential, log-normal, log-logistic and Gompertz).

P-values for the coefficients of the simplified equations for CV-death for all considered distributions are provided in "*B1_Risk equations_Appendix 1".* This also includes p-values for simplified equations for other outcomes and for equations fitted to subgroups of interest.

B3. Priority question. The ERG disagrees with the company's assessment of proportional hazards (PHs) between the empagliflozin and placebo CV-death KM data in EMPEROR-Reduced. Furthermore, the current joint modelling approach reflects a poor representation of the underlying KM data for CV-death in the EMPEROR-Reduced trial. Therefore, please provide an option in the economic model where CV-death is independently modelled for the empagliflozin and comparator arms.

Please note that while the number of patients at risk at the end of the KM curve are small, there are about 50% of patients at risk at 16 months where the KM curves overlap/cross.

The rationale for joint modelling for CVM and use of this as the base-case in the CE model is based on the following considerations. We elaborate further in B3 a-c).

- CV deaths were relatively rare with around 10% of patients dying due to a CV event. With low event counts, the observed shape of the KM curves must be interpreted carefully to distinguish between chance variation and true signals of change. For instance, while the curves for the two arms connect around month 15, they separate again thereafter with a larger separation until about month 25. The curves cross after that point, but at this stage only 15% of the original population is still at risk; this leads to long flat periods where no events are observed or curves drop sharply due to a single CV-death, both of which contribute to the curves crossing near the end of the observation period. Thus, we interpret the overall pattern as indicative of a small treatment benefit, and the data not having sufficient event counts to be able to reliably detect and model any underlying non-proportionality.
- The CV-death equations include current KCCQ-CSS quartile as a time-varying predictor, which captures at least part of effect of treatment on CV-death overtime. The time-dependent nature of KCCQ-CSS in the equation indirectly captures fluctuations in treatment effect over time.
- We considered separate fits for the two arms. These are discussed in Question B3b. These fits showed implausible long-term patterns with projections for placebo leading the longer mean and maximum time to CV-death with all the tested distributions. Thus, separate fits would lead to

divergent projections, which is not clinically plausible and could not be used in that form in the CE model. Limitations with separately fitted distributions are further discussed in responses to a) and b).

- A sensitivity analysis was conducted in the CE model where no treatment benefit is applied for empagliflozin for CV-death, which had a minor impact on the ICER (increase of less than £350/LYG), indicating that non-proportionality is a minor consideration for the committee.

While conducting the analysis please:

a) Justify any assumptions made to fit the data at the end of the KM curves due to low numbers of patients at risk.

As noted above, equations fitted to the two treatment arms were not incorporated in the CE model, but these fits were considered and reported in **Table** *27* and Appendix 2.

b) Provide the rationale for using a particular survival model as currently done for the joint model (i.e. AIC and BIC criteria; visual inspection of the data; clinical plausibility).

As most of the observed deaths in the trial are CV-related, considerations for selecting an optimal fit apply to all-cause death. Diagnostic plots for the tested models are shown (Table 27 and
Figure 10) (Source: Appendix 2, Section 4.3.2.1). These suggest comparable fits for Weibull, Gompertz and the log-logistic distributions in general, and a good fit for the exponential distribution when fitted to the placebo curve alone but not in the empagliflozin arm or when the data are fitted jointly with treatment as a predictor. The log-normal and generalised gamma distribution had slightly weaker fit.

Predicted curves from these fitted models show close fit to the observed data during the observation and start to separate near the tail (
Figure 10) thus, differences in long-term projections was used to determine the best fitting distribution.

Table *28* summarises the predicted mean (life expectancy) and maximum survival time from each of the fitted distributions, and Figure 11 shows the shape of the long-term projections. As these fits are for death for a specific cause, considerations of the shape of the tail differ from those for all-cause mortality. In particular, very long maximum survival times are not necessarily indicative of poor projection, as the risk from a specific cause would be expected to decline over time.

The jointly fitted Weibull distribution was selected based on the following considerations:

- Arm-specific fits for all except the exponential fit produce estimates of longer survival with placebo. This is due to the observed curves cross and yield predicted curves that will also cross and produce shorter survival with empagliflozin. As noted above, the crossing of the observed curves appears to be likely due to the frequency and timing of observed events. Thus, only joint fits were considered further.
- Similarly, the jointly fitted gamma distribution yield later CV-death times for placebo, which is inconsistent with the results observed in the trial.
- While the mean time to CV-related death from the exponential were similar to those from the Weibull fits, it seems plausible that the risk of CV events and death as a consequence likely increases over time rather than remain constant as assumed with the exponential distribution.
- The log-normal and log-logistic distributions show improbably long predicted mean time-to-CV death.
- The Gompertz fit yields shorter expected times and a sharply increasing hazard (dropping curve) of CV-related death suggesting a mean event time of five to seven years, which may exaggerate the risk.

Table 27. Goodness-of-fit Statistics (AICC/BIC) for Tested Distributions for CV-death

Model	Fitting to Placebo Alone		Fitting to Empagliflozin 10 mg Alone		Joint Fitting (Treatment as Predictor)	
Distribution	AICC	BIC	AICC	BIC	AICC	BIC
Exponential	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Weibull	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Log-normal	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Log-logistic	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Generalised gamma	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: AICC = corrected Akaike's Information Criterion; BIC = Bayesian Information Criterion

Figure 10. Observed and Predicted Time to CV-death Distribution

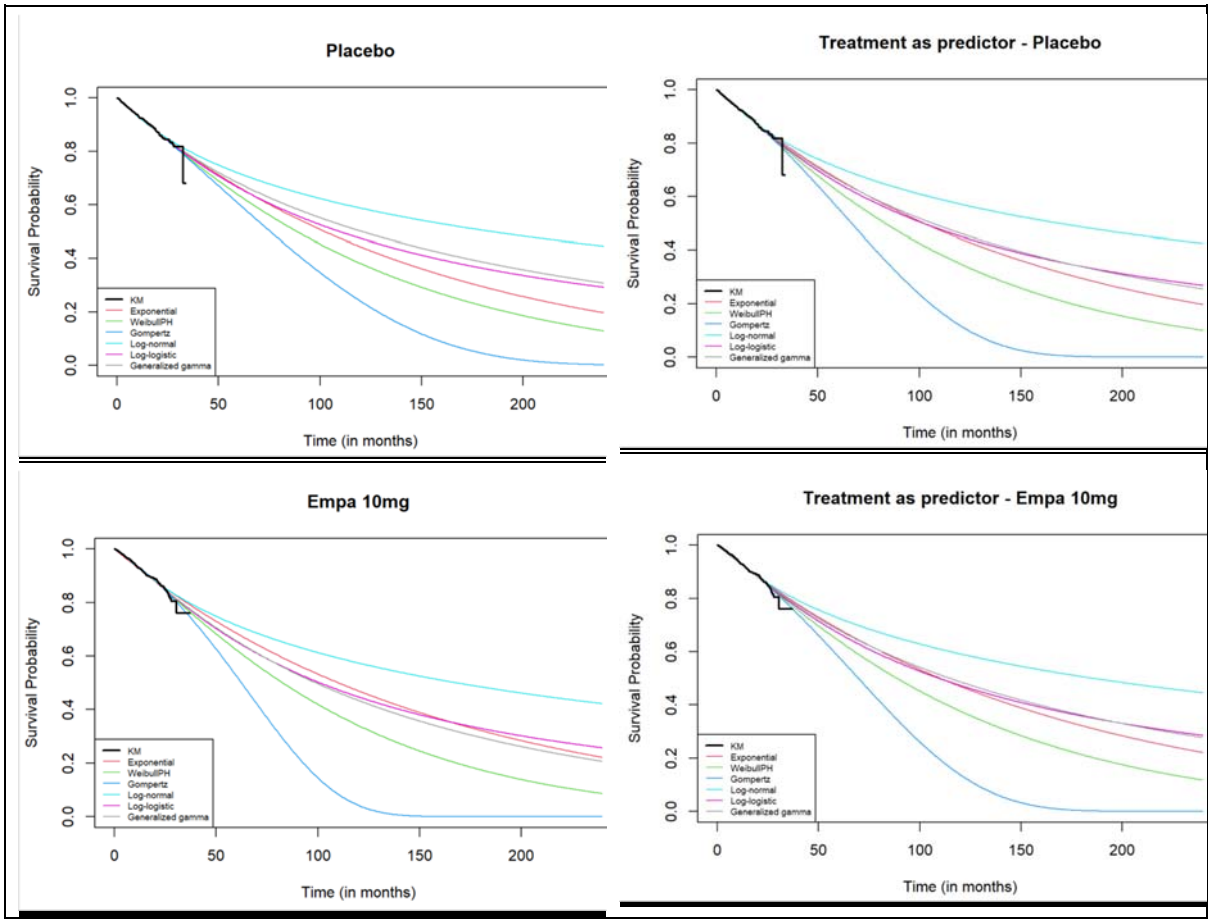
Abbreviation: empa = empagliflozin

Table 28. Mean and Maximum Survival (in months) from Fitted Distributions for CV-death

Treatment	Placebo				Empagliflozin 10 mg			
Model	Fitting to Placebo Alone		Joint Fitting (Treatment as Predictor)		Fitting to Empagliflozin 10 mg Alone		Joint Fitting (Treatment as Predictor)	
Predictions (month)	Mean	Max*	Mean	Max	Mean	Max	Mean	Max
Exponential	[redacted]							
Weibull	[redacted]							
Gompertz	[redacted]							
Log-normal	[redacted]							
Log-logistic	[redacted]							
Generalised gamma	[redacted]							

* 99% quantile from the fitted distribution – time after which only 1% of patients are alive.

Figure 11. Long-term Projections of CV-related Death



Abbreviation: empa = empagliflozin

c) Provide an option in the economic model where alternative distributions (fitted independently to each treatment arm) can be selected to run the analysis)

As stated in B2a, we do not implement separate fits into the CE model. Instead, we allow the user flexibility to relax the proportional hazards assumption for CE model by setting the coefficient for treatment to zero from a set point in time specified by the user. Turning treatment off at a certain time point is more conservative than implementing independent fits for treatment and placebo. Even this more conservative assumption has a limited impact on the ICER. For instance, turning off treatment effect at month 60 in the mortality equations changes the ICER from £4,717/QALY to £4,831/QALY. The conclusion from these analyses is that the proportional hazards assumption is only a minor consideration for the committee to base their recommendation for empagliflozin.

B4. Priority question. Include a tab in the Excel model where all the alternative survival distributions from B3 are plotted against the KM data on CV-death from EMPEROR-Reduced using the independent modelling approach. Please ensure the KM data and fitted survival data (underpinning the plots) are provided in the same tab.

The alternative survival distributions suggested in B3 are not implemented in the model due to the reasons explained in our response to B3 a,b,c. Hence, the requested curves for KM data versus fitted survival have not been added into the model.

B5. Priority question. The ERG's clinical experts advised that the population in the trial is around a decade younger than the average age of HF patients in clinical practice and that this is likely to have an impact on the outcomes of the model. The view of the ERG is that there are two important populations available for analysis: a 'trial only' analysis of a younger, higher risk population from EMPEROR-Reduced and a 'real-world' population representing an older baseline age as reflected in PULSE. The ERG accepts that these analyses will have limitations but ask that the company highlight those limitations in their response.

The ERG's clinical experts advised that the mean age of the population likely to be treated with empagliflozin in the UK is expected to be around 10 years older than the average patient enrolled in the EMPEROR-Reduced trial (mean age: 66.8 years). This is also supported by the PULSE study where the mean age is [redacted] years (13).

To account for this and reflect the average age of patients expected to be treated in the UK, a subgroup analysis was performed. This included only the patients in EMPRERROR-Reduced who were 65 years or older. The average age of this older subgroup is 73.8 years, and therefore it is reasonable to assume that this subgroup closely reflects the average age of the target population. All risk equations were updated for this subgroup in the CE model. Furthermore, a more `pessimistic' survival distribution and (specifically the Gompertz) was used for all-cause and CV-death, as requested by the ERG.

The results of the subgroup analysis are provided in Table 29. Overall, the ICER increased from [redacted] (after amendments implemented as part of B15 & B19) to [redacted] remaining well below the £20,000/QALY willingness to pay threshold.

Table 29: Deterministic cost-effectiveness of empagliflozin as an add-on to standard care in the ITT population and the ≥ 65 years subgroup with a Gompertz distribution for all-cause and CV-death.

Technology	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
ITT population: Base-case results in the original submission							
SoC	£16,887	[redacted]	[redacted]	-	-	-	-
Empagliflozin + SoC	£17,950	[redacted]	[redacted]	£1,063	[redacted]	[redacted]	[redacted]
≥ 65 years subgroup: Updated results in response to B5							
SoC	£11,638	[redacted]	[redacted]				
Empagliflozin + SoC	£12,635	[redacted]	[redacted]	£998	[redacted]	[redacted]	[redacted]

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life-years gained; SoC, standard of care; QALYs, quality-adjusted life year. a. CE model selections: Age≥65 years, Gompertz for survival

It should be noted, however, that although this subgroup analysis pertains to a population with a similar average age as that expected to be treated in the UK, it is a `higher-risk' population than PULSE. This is because the EMPEROR-Reduced trial was enriched with patients with a more advanced disease who were also at higher risk of HHF, evident in the lower mean ejection fraction (27.2 vs 32.1 in EMPEROR-Reduced placebo arm and PULSE, respectively) and lower median NT-proBNP (1926 vs 1023 EMPEROR-Reduced PBO arm and PULSE, respectively) (7, 11).

The conclusion from this scenario is that age is a minor consideration for the committee. Even when we more closely approximate the age of patients in the PULSE study, empagliflozin still offers value for money to the NHS, with the ICER still being below <£20,000/QALY.

Please provide a version of the model that uses the trial outcomes for the high-risk population, with the EMPEROR-Reduced hospitalisation rate but a more "pessimistic" survival curve (e.g. the Gompertz).

This scenario has been implemented in CE model. Please see Appendix 3.

B6. Priority Question. Please provide a version of the model that captures the 'real-world' population as expected in clinical practice with an average age in line with the PULSE study. The ERG is aware that the company has selected the Weibull extrapolation for CV-related deaths and validated this using the PULSE study. This survival curve could be considered reflective of the "real-world" population discussed in B5. The ERG suggests the company performs a similar analysis for HHF, again using the PULSE study to validate the results. The ERG accepts that these analyses will have limitations but ask that the company highlight those limitations in their response.

To reflect a population like the PULSE study, the ≥65 years subgroup from EMPEROR-Reduced was used to predict LYs, QALYs and an ICER in the CE model. This is similar to the approach in B5; however, a Weibull distribution was selected instead of a Gompertz for mortality. This approach is reasonable because the average age in the ≥65 years subgroup and PULSE was similar (73.8 years across in EMPEROR-Reduced vs 72.2 years in PULSE).

We believe that this is the best approach to approximate an older population in the CE model that is more reflective of the UK HFrEF population.

The alternative to this approach -utilizing the PULSE data directly into the CE model as the SoC arm - is not feasible. This is because KCCQ-CSS is not routinely recorded in the HES or CPRD databases. Therefore, HHF rates cannot be derived separately per KCCQ-CSS health state as required for the CE model. Furthermore, even if the model's granularity was reduced to aggregate all KCCQ-CSS health states together, and hence ignore differences in disease severity and consequently in outcomes, it would still not be feasible to obtain HHF rate estimates for both treatment arms using the PULSE data without accepting significant uncertainty. A match adjusted indirect comparison (MAIC) would be required, however this relies on there being sufficient overlap in study populations. This assumption is unlikely to be satisfied given the disparity in age between the two data sets, and the specific inclusion criteria in EMPEROR requiring patients to have specific NT-proBNP for

each ejection fraction cut-off. The ≥65 group represents only one third of the trial population, reducing the effective sample size significantly (7).

This approach of approximating an older population with a subgroup analysis has limitations. The ≥65 years subgroup does not reflect the HHF rate observed in UK population. The HHF rate in the ≥65 years group in EMPEROR-Reduced (20.2 events per 100 PYs, assuming the baseline KCCQ-CSS distribution in the trial) was higher than in the PULSE study [redacted]) events per 100 PYs [age and sex adjusted]). The difference in these rates might be explained by the limited accuracy of HHF recording in HES. As stated in A7, non-HF hospitalisation cannot be clearly separated from HF hospitalisation. Unlike in the real-world, HHF in EMPEROR-Reduced were adjudicated by committee according to a strict protocol. In the real world, an elderly patient might be admitted to wards other than cardiology, and therefore HHF may not be recorded as the primary reason for hospitalisation because general physicians and other specialists may not recognise the symptoms of acute HF.

If we accept these limitations, the results of this analysis illustrate that empagliflozin + SoC remains highly cost-effective (ICER: £[redacted]/QALY gained) (i.e. <£20,000/QALY) in this older population subgroup that more closely reflects the `real-world' population (Table 31). Therefore, generalizability to the 'real world' population is only a minor consideration for the committee.

Table 30. Baseline characteristics of PULSE HFrEF cohort (incident and prevalent) and EMPEROR-Reduced placebo arm of subgroup of patients with an age ≥ 65 years.

Patients characteristics	PULSE (n=68,780)	EMPEROR PBO Age ≥65 years subgroup (n=1,127)
Age at index (years)	[redacted]	73.8
Sex, females (%)	[redacted]	26.3%
Hospitalisation for HF in prior 12 months	[redacted]	29.2%
BMI (kg/m^2), mean	[redacted]	27.3
LVEF, mean	[redacted]	28.0

NT-ProBNP, mean	[redacted]	2,189.1
Systolic Blood Pressure, mean	[redacted]	122.6
Heart Rate (bpm), mean	[redacted]	70.4
eGFR (ml/min/1.73m^2)	[redacted]	55.1
Type 2 diabetes (%)	[redacted]	49.2%

*Over 95% missing data in PULSE.
Abbreviations: BMI, body mass index; bpm, beats per minute; HF, heart failure; PBO, Placebo; LVEF, Left-Ventricular Ejection Fraction; NT-proBNP, N-terminal pro-brain natriuretic peptide; eGFR, estimated glomerular filtration rate;

Table 31: Deterministic cost-effectiveness of empagliflozin as an add-on to standard care in the ITT population and the >= 65 years subgroup[a]

Technology	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
ITT population: Base-case results in the original submission							
SoC	£16,887	[redacted]	[redacted]	-	-	-	-
Empagliflozin + SoC	£17,950	[redacted]	[redacted]	£1,063	[redacted]	[redacted]	[redacted]
≥ 65 years subgroup: Updated results in response to B6							
SoC	£15,198	[redacted]	[redacted]				
Empagliflozin + SoC	£16,436	[redacted]	[redacted]	£1,238	[redacted]	[redacted]	[redacted]

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life-years gained; QALYs, quality-adjusted life year; SoC, standard of care.
CE model setting selections: Age≥65 years, Weibull for survival

CV Hospitalisations

B7. Priority question. Please provide the p-values associated with [redacted]

P-values for the coefficients of the simplified equations for HHF are provided in the attached spreadsheet (Appendix 1).

B8. Priority question. The ERG disagrees with the company's assessment that the [redacted]

Therefore, can the company please consider an alternative modelling approach to [redacted].

The equation for HHF does not explicitly include a term for time but it should be noted that since KCCQ-CSS is included as a time-varying predictor, the rates do vary over time based on changes in KCCQ-CSS levels.

The plot referenced above (Evidera 2021,Figure 10) does show fluctuations in gap between the rates in the two arms, but these must be interpreted carefully since the rates in a given month may be based on a few events [redacted] in some cases), and the number of patients contributing to this calculation declines over time. Therefore, observed variations may not necessarily indicate a real change in the rate between the arms. For example, the curves for placebo and empagliflozin at month 12 cross but the rate increases sharply for placebo in the next month, exceeding the rate observed in immediately previous months and separating the curves again. Thus, the convergence at month 12 is likely spurious and may not indicate a real change in

the risk. Similarly, the curves connect at month 17, but separate again in the next month. The curves cross permanently after month 21, but about 25% of the original sample are still followed at this time and less than 15% are still followed by month 25 (see numbers at risk in Figure 2 of Appendix 2). Thus, the drop in HHF rate in the later part of the curves and their crossing may be due to high variability expected with low patient counts rather than a true signal of a change.

We examined the role of time in the simplified KCCQ-CSS-based equation for HHF in the ITT population by adding time and an alternative model allowing an interaction between time and treatment. Coefficients are reported in

Table ***6*** and show a weak negative slope for the time (in months) with a p-value [redacted]. The negative slope implies a declining rate of hospitalisation over time, which is not clinically plausible over the longer term that would be projected in the CE model. The interaction term added to this model was not statistically significant (p-value = 0.40) and suggests an implausible long-term pattern with rates declining for placebo (log-rate changing by -[redacted] per month), while the slope in the empagliflozin arm is nearly flat (-0[redacted]). This leads to divergence over time with lower rates for placebo, which is not plausible (Table 32).

We believe the patterns captured in these equations are affected by low patient counts near the end of follow-up and do not represent clinically plausible projection patterns to be used in the CE model. UK clinical experts agreed that the rate of HHF does not decrease over time.

Rather than adding alternative equations to the CE model, BI have added an option that allows the user to turn off the treatment effect for HHF from a set time onwards. With the added option, the timing of this can be varied. For instance, turning off the effect as of month 60 in the HHF equation changes the ICER from [redacted]/QALY to [redacted]/QALY.

Therefore, assumptions about varying rate of hospitalisation over time has limited impact on the ICER, and therefore is only a minor consideration for the committee.

Table 32. Risk equations for HHF with time effect and interaction term

	Estimate	SE	P-Value
Simplified KCCQ-CSS-based Equation for ITT Population with Common Time Effect			
Intercept	[redacted]	[redacted]	[redacted]
TRT: Empa 10mg	[redacted]	[redacted]	[redacted]
KCCQ-CSS [55.2,75]	[redacted]	[redacted]	[redacted]
KCCQ-CSS [75,89.6]	[redacted]	[redacted]	[redacted]
KCCQ-CSS [89.6,100]	[redacted]	[redacted]	[redacted]
Time (in Months)	[redacted]	[redacted]	[redacted]
Simplified KCCQ-CSS-based Equation for ITT Population with Time Effect and Interaction with Treatment			
Intercept	[redacted]	[redacted]	[redacted]
TRT: Empa 10mg	[redacted]	[redacted]	[redacted]
KCCQ-CSS [55.2,75]	[redacted]	[redacted]	[redacted]
KCCQ-CSS [75,89.6]	[redacted]	[redacted]	[redacted]
KCCQ-CSS [89.6,100]	[redacted]	[redacted]	[redacted]
Time (in Months)	[redacted]	[redacted]	[redacted]
TRT:Empa 10mg x Time (in Months)	[redacted]	[redacted]	[redacted]

B9. Priority question. Please provide the number of hospitalisation events predicted by the economic model in both arms and compare these estimates to the number of hospitalisations in EMPEROR-Reduced (for the relevant time period).

The median follow-up duration in the EMPEROR-Reduced trial was 16 months. The total number of HHF recorded during the trial, and that predicted by the cost-effectiveness model in the cycles 0 to 16 is shown in Table . The model predicts fewer hospitalisations of patients receiving empagliflozin + SoC compared to those receiving only SoC, in accordance with the observed hospitalisation rates in the EMPEROR-Reduced trial. At the end of the first sixteen model cycles, patients receiving empagliflozin are predicted to experience a 23% reduction (i.e., 374/485) in the number of HHF episodes compared to those on SoC, which is consistent with the 28% reduction (i.e., 246/342) seen in the trial and is a conservative estimate. The difference in the predicted event rates is almost identical to that observed in the trial (i.e., [redacted] in the model versus [redacted] in the trial) (**Table *33***).

In conclusion, this analysis indicates that the CE model has good face validity when compared to the trial results, and therefore this is a minor consideration for the committee.

Table 33. The total number of hospitalisations for HF recorded in the EMPEROR-Reduced and predicted by the model in 16 months of patient follow-up

	EMPEROR-Reduced			CE model		
	Empagliflozin + SoC (N=1863)	Placebo +SoC (N=1867)	Difference	Empagliflozin + SoC (output per 1863 patients)	Placebo +SoC (output per 1867 patients)	Difference
Total No of HHF during the 16 months of follow-up	246	342	-96	[redacted]	[redacted]	[redacted]
Total HHF event rate (events/100 patient-yr)	10.7	15.5	-4.8	[redacted]	[redacted]	[redacted]

Abbreviations: HHF, hospitalisation for heart failure; SoC, standard of care.

B10. Priority question. There appears to be a [redacted] for empagliflozin + SoC than SoC alone reported in the company submission (as covered in questions [redacted] and [redacted]). The ERG's clinical experts advised that this may be due to [redacted]. If the answer to [redacted], please account for this appropriately in the model.

This question is linked to question A7 and A8. As noted in **Table** 10, the proportion of patients in the empagliflozin arm experiencing a non-HHF was slightly lower than in the placebo arm. Therefore, it is unlikely that incorporating non-HHF events into the CE model will have an impact on the ICER and has not been incorporated. This

is a conservative assumption. Based on these data, the rates of non-HHF is only a minor consideration for the committee.

Transition probabilities

B11. Priority question. The hazard ratios (HRs) provided in Figure 20 of the company submission (page 135) and in the EMPEROR-Reduced clinical study report, Table 15.2.3.6: 5) which report the effect of empagliflozin vs placebo on mean KCCQ-CSS over time seem to match Figure 15.2.3.6: 4 in the EMPEROR-Reduced clinical study report (page 465). However, Figure 15.2.3.6: 4 differs from Figure 20 in the company submission. Please explain the difference in these figures and point the ERG to the figure corresponding to the HRs reported in the clinical study report.

The difference between Figure 20 in the company submission and Figure 15.2.3.6:4 in the EMPEROR-Reduced clinical study report (11) is in the data set from which they have been derived. Figure 20 (and CTR, Table 15.2.3.6:5) pertains to all observed data whether on- or off-treatment [i.e. RS(OC-AD), randomised set with observed cases including data after treatment discontinuation] with no imputation for deaths. Figure 15.2.3.6:4 (and Table 15.2.3.6:3 on the page 461 of the CTR) on the other hand pertains to the observed cases on-treatment only [TS (OC-OT)]. Also Figure 15.2.3.6:4 plots the absolute value of the mean KCCQ-CSS score over time, while Figure 20 in the company submission plots the mean difference in the KCCQ-CSS from baseline over time. The adjusted mean difference in KCCQ-CSS score between empagliflozin and placebo at different time points shown in Figure 20 match those shown in Table 15.2.3.6:5 of the CTR but *do not* match those in Figure 15.2.3.6.4 (or Table 15.2.3.6:3). Data from Table 15.2.3.6:5 is also shown in Table 8 of these clarification questions.

These differences in data sources are shown in Table 34.

Table 34. Clarification of data sources and assumptions for KCCQ-CSS

Source, and data set	Data set	Adjusted mean difference, empagliflozin vs placebo (95% CI)		
		12 weeks	32 weeks	52 weeks
Figure 20 from the company submission	RS (OC-AD)	1.94 (0.96-2.93)	1.35 (0.28-2.42)	1.61 (0.39-2.84)
Table 15.2.3.6:5 (CSR, p470),		1.94 (0.96-2.93)	1.35 (0.28-2.42)	1.61 (0.39-2.84)
Figure 15.2.3.6:4 (CSR, p465), TS (OC-OT)	TS (OC-OT)	NR	NR	NR
Table 15.2.3.6:3 (CSR, p461), TS (OC-OT)		[redacted]	[redacted]	[redacted]

Abbreviations: RS, randomised set; TS, treated set; OC-AD, observed case including data after treatment discontinuation; OC-OT, observed case on-treatment.

B12. Priority question. Please provide the overall change in mean KCCQ-CSS estimated in the model (i.e. considering the baseline mean KCCQ-CSS) for month 3; month 8; and month 12. Please compare the estimated mean changes for empagliflozin and SoC with those observed in EMPEROR-Reduced for the same time points.

The model only tracks the change in proportion of patients in each KCCQ-CSS quartile and cannot report on the overall change in mean KCCQ-CSS over time. Table 35 below compares the proportion of patients observed in each KCCQ-CSS quartile at weeks 12, 32, and 52 (with imputation) as reported in EMPEROR-Reduced versus the model reported proportions at months 3, 8, and 12 (i.e., closest matching cycle) along with the difference in each reported outcome).

These results show that the observed and predicted KCCQ-CSS transitions are closely matched, and therefore this a minor consideration for the committee.

Table 35. Proportion of patients in different KCCQ-CSS quartiles over time, trial versus model predicted

Trial	Visit Time (week)	Empagliflozin + SoC				SoC			
		KCCQ-CSS Quartile 1 0 to55.2 (%)	KCCQ-CSS Quartile 2 55.2 to 75 (%)	KCCQ-CSS Quartile 3 75 to 89.6 (%)	KCCQ-CSS Quartile 4 89.6 to 100 (%)	KCCQ-CSS Quartile 1 0 to 55.2 (%)	KCCQ-CSS Quartile 2 55.2 to 75 (%)	KCCQ-CSS Quartile 3 75 to 89.6 (%)	KCCQ-CSS Quartile 4 89.6 to 100 (%)
	Baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	32	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	52	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Model Predicted	**Model Cycle (month)**	**Empagliflozin + SoC**				**SoC**			
		KCCQ-CSS Quartile 1 0 to 55.2 (%)	KCCQ-CSS Quartile 2 55.2 to 75 (%)	KCCQ-CSS Quartile 3 75 to 89.6 (%)	KCCQ-CSS Quartile 4 89.6 to 100 (%)	KCCQ-CSS Quartile 1 0 to 55.2 (%)	KCCQ-CSS Quartile 2 55.2 to 75 (%)	KCCQ-CSS Quartile 3 75 to 89.6 (%)	KCCQ-CSS Quartile 4 89.6 to 100 (%)
	Baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	8	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Difference Between Trial and Model Predicted	**Model Cycle (month)**	**Empagliflozin + SoC**				**SoC**			
		KCCQ-CSS Quartile 1 0 to 55.2 (%)	KCCQ-CSS Quartile 2 55.2 to 75 (%)	KCCQ-CSS Quartile 3 75 to 89.6 (%)	KCCQ-CSS Quartile 4 89.6 to 100 (%)	KCCQ-CSS Quartile 1 0 to 55.2 (%)	KCCQ-CSS Quartile 2 55.2 to 75 (%)	KCCQ-CSS Quartile 3 75 to 89.6 (%)	KCCQ-CSS Quartile 4 89.6 to 100 (%)
	Baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	8	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Treatment discontinuation

B13. Priority question. Please include a tab in the Excel model where all the alternative survival distributions provided in the economic model to estimate time-to-treatment discontinuation (TTD) are plotted against the respective TTD KM data from EMPEROR-Reduced. Please ensure the KM data and fitted survival data (underpinning the plots) are provided in the same tab.

A new tab (i.e., KM Curves - TTD) has been added into the CE model where the selected survival distribution for TTD is plotted against the KM data over time. Alternative distributions can be readily evaluated by manipulating the dropdown menu in F144 in the "Clinical Inputs" tab, analogous to the response to B1 above. The updated CE model is Appendix 3.

B14. Priority question. Please explain [redacted] Please provide the p-values for all the different distributions considered (exponential, Weibull, log-normal, log-logistic and Gompertz).

There was no specific reason. P-values for the coefficients of the simplified equations for TTD for all considered distributions are provided in Appendix 1. This also includes p-values for simplified equations for other outcomes and for equations fitted to subgroups of interest.

B15. Priority question. The ERG is unsure why TTD data from EMPEROR-Reduced was jointly fitted between treatment arms. Given the nature of TTD data it would seem logical that a single-arm parametric curve was fitted to trial data on discontinuation with empagliflozin. Please:

a) **Justify the decision of fitting a joint model.**

Joint modelling was considered because the observed TTD in the two arms was similar, suggesting that a common distribution may adequately fit the data. This allows leveraging all the available data to estimate the parameters of the fitted

equations. We agree, however, that fitting equations to the empagliflozin arm aligns more closely with the way the equation is used in the CE model, so the base-case was modified to use the single-arm equation.

Upon review of TTD analyses, it was noted that some patients who were recorded as having discontinued treatment with a fatal AE as the reason. These patients were originally censored to avoid double-counting discontinuations as death in the CE model already would trigger stopping treatment. Some of the deaths of patients who discontinue treatment with fatal AE occur sometime after the recorded discontinuation time – that is, patients were alive at the time of discontinuation and die sometime later. The analyses were rerun with a revised TTD variable where discontinuations with fatal AE are only censored when death occurs on the same day as discontinuation. This effectively increases the number of discontinuations in the analyses. Parametric fitting assessment led to similar conclusions with Weibull showing best statistical fit statistics and exponential fitting the later part of the observed curves more closely. The Statistical Analysis report is updated with the revised analyses (provided as Appendix 2), and the exponential distribution fitted to the empagliflozin arm included in the CE model as base-case. Using the Weibull distribution as an alternative scenario led to a very similar ICER (less than £50 difference).

b) Provide a scenario analysis where one TTD curve is fitted to the TTD KM data for empagliflozin from EMPEROR-Reduced.

As noted in a), the base-case was modified to use the empagliflozin-based equation based on an exponential distribution. The cost-effectiveness with this scenario is £[redacted].]

B16. Priority question. The ERG notes that the exponential model used to fit TTD data in the model yields the highest AIC and BIC statistics. Additionally, the use of the exponential model implies a constant rate for treatment discontinuation over time. Please:

a) **Provide any long-term data available for empagliflozin (or alternatively dapagliflozin) to justify the long-term predictions around treatment discontinuation with this class of drugs;**

   BI is not aware of any long-term data exists for treatment discontinuation for empagliflozin or dapagliflozin in HF. It should be noted that the assumptions made in the CE model for treatment discontinuation are conservative, i.e. patients receive the costs and benefits of SoC following treatment discontinuation.

   Assuming no treatment discontinuation for empagliflozin results in an ICER of £[redacted]QALY, which is still below the £20,000/QALY threshold. Thus, assumptions made about treatment discontinuation is only a minor consideration for the committee.

b) **Provide a scenario analysis where the best fitting distribution (Weibull) is used in the model.**

   Modifying the parametric distribution for TTD from exponential to Weibull does not have a meaningful impact on ICER for empagliflozin as an add-on to SoC, as shown in Table . Therefore, the choice of distribution for TTD is only a minor consideration for the committee.

Table 36. Deterministic cost-effectiveness of empagliflozin as an add-on to standard care, with the resource use for disease management as shown in Table A

Technology	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
Assuming exponential distribution for TTD							
SoC	£16,911	[redacted]	[redacted]				
Empagliflozin + SoC	£17,837	[redacted]	[redacted]	£926	[redacted]	[redacted]	[redacted]
Assuming Weibull distribution for TTD							
SoC	£16,911	[redacted]					
Empagliflozin + SoC	£17,893	[redacted]		£981	[redacted]	[redacted]	[redacted]

c) Abbreviations: LYG, life-years gained; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; SoC, standard of care; TTD, time-to-treatment discontinuation.

B17. The committee in TA679 preferred the approach using standard statistical quartiles of the KCCQ-CSS scale (0-<25, 25-<50, 50-<75, 75-100) as health states in the model. Please add the function to use these quartiles into the empagliflozin model.

BI have confirmed with the ERG that it is not necessary to implement standard quartiles (e.g. KCCQ-CSS 0-25, 26-50, 51-75) in the cost-effectiveness model for empagliflozin. This is because standard quartiles were not used in the DAPA-HF model for TA679(4). The quartiles used in the DAPA-HF model were KCCQ-TSS 0 to <58, 58 to <77, 77 to <92 and 92 to 100. The quartile cut-offs used in the empagliflozin submission were similar (KCCQ-CSS: 0 to <55.2, 55.2 to <75, 75to <89.6 and 89.6 to 100). In the dapagliflozin FAD, the ERG noted that "*cut-offs for the quartiles chosen by the company to measure KCCQ-TSS in the model were arbitrary. But it said it expected that using other cut-offs or approaches to grouping would minimally affect the cost-effectiveness results. The committee concluded that the company's model structure was appropriate for decision making*"(4). Given the similarities in approach, we anticipate similar deliberations for this appraisal.

B18. Priority Question. Table 20 in the Company submission states that 0% of either arm has a sustained eGFR rate of <15ml/min or <10ml/min, however in the model there is an assumed rate of 88% of the patients who enter the renal model have a sustained eGFR reduction. Please explain the rationale for this number.

Table 20 of the company submission (which corresponds to the Table 11.1.2.6:1 of the Clinical Study Report), shows that the composite renal endpoint was experienced by 88 patients in the EMPEROR-Reduced trial (58 patients in the placebo arm and 30 patients in the empagliflozin arm). These patients experienced only two of the four types of adverse renal outcomes included in the definition of the composite renal outcome (listed in Table 37). While no patients received the renal transplant or had sustained eGFR <15 mL/min/1.73m^2 (for those with baseline eGFR ≥30 mL/min/1.73m^2) or sustained eGFR <10 mL/min/1.73m^2 (for those with baseline eGFR <30mL/min/1.73m^2), there were [redacted] patients on chronic dialysis and [redacted] patients with sustained eGFR reduction from baseline of ≥ 40% during the trial follow-up. Proportionally, therefore, chronic dialysis comprised [redacted]) of all serious renal outcomes, while sustained reduction in eGFR from baseline of ≥40% comprised the remaining [redacted] of renal events. These proportions were then used to derive the weighted average cost of the composite renal outcome as experienced by subjects in the EMPEROR-Reduced trial (Table 37).

Table 37. Cox regression analysis of time to first renal event in the randomised set

Time to composite renal outcome	Placebo (N=1867)	Empagliflozin (N=1863)
Patients with the composite renal endpoint, N (%)	58 (3.1)	30 (1.6)
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

Source: EMPEROR-Reduced CSR, Table 11.1.2.6:1
Abbreviations: eGFR, estimated glomerular filtration rate

B19. Priority Question. Please remove the half-cycle correction applied to the discounting factor on costs and benefits (i.e. in cycle 0 of the model the discounting factor should be 1 and not 0.999, etc.).

Thank you for pointing this out. The half-cycle correction that was applied to the discounting factor on costs and benefits has now been removed. This modification did not lead to any meaningful change in the ICER; the estimated cost per HHF avoided was the only parameter impacted by this correction (from [redacted] to [redacted]). Importantly, this change was also applied in all the other scenarios implemented as part of our responses to the points for clarification.

Utility data

B20. Priority question. Please include age-related utility decrements throughout the model time horizon using the algorithm published by Ara and Brazier 2010 (please note that at 25 years there are still 1% of patients alive in the model).

A scenario was programmed into the model to allow for an age-adjustment to KCCQ-CSS quartile utility values over time based on UK general population. A multiplier was calculated based on cohort age and sex using the formula for general population EQ-5D reported in Ara and Brazier (2000) (65), like the age-adjustment considered by the ERG in the NICE process. The multiplier was incorporated into the utility calculations in the model engine sheets for empagliflozin + SoC and SoC. In the alternative analysis applying the age adjusted utilities, the ICER is £[redacted]/QALY. This very slight increase in the ICER compared to the base-case (no age adjusted utilities) is driven by the slightly longer survival of patients in the empagliflozin + SoC arm compared with those on SoC. This scenario analysis has limited impact on the ICER and therefore is a minor consideration for the committee.

B21. Priority question. In the company submission (page 152) it is stated that, "as the trial-derived utility value for KCCQ-CSS quartile 4 ([redacted]) was higher than the utility of UK general population aged 60 to 69 years (0.7740) reported by Sullivan et al. 2011 an age-adjustment was applied. Under this adjustment, utility values for KCCQ-CSS quartile 1–3 were reduced by the relative difference between EMPEROR-Reduced observed utility for KCCQ-CSS quartile 4 and published utility of UK general population aged 60 to 69 years". However, the adjustment applied for the 1–3 KCCQ-CSS quartiles was based on the absolute difference between the EMPEROR-Reduced utility and the Sullivan value (0.7740). Please replace the [redacted]difference ([redacted] minus 0.7740) adjustment applied in the three quantiles by the relative difference ([redacted]%) in the utility values.

We have implemented the suggested approach using the relative difference in the utility values instead of the absolute difference. The adjusted utility values changed from [redacted], [redacted], [redacted], [redacted] for KCCQ-CSS Q1 to Q4, respectively, to [redacted], [redacted], [redacted], [redacted]. As a result of this change, the estimated overall QALYs for the empagliflozin + SoC and the SoC arms decreased from [redacted] and [redacted] to [redacted] and [redacted], respectively. The difference in QALYs between the two arms remained at [redacted] (a difference was observed only on the third decimal), and the ICER changed from [redacted] to [redacted]. Overall, there was no meaningful impact on cost-effectiveness, and therefore this adjustment is only a minor consideration for the committee.

B22. Priority question. Please clarify if diabetes at baseline was investigated as an explanatory variable for changes in patients' quality of life and provide a justification for why this variable was not included in the regression models.

Diabetes status at baseline was tested in the utility equation but was not retained as its effect was not statistically significant.

B23. Priority question. Please explain why the baseline utility value from EMPEROR-Reduced used in the regression models in the model is zero (and correct this baseline estimate if this was a mistake in inputting the value).

Baseline utility in the utility equation is standardized; meaning the observed values are subtracted from the mean and divided by the SD. Thus, a value of zero for baseline utility corresponds to setting it to the mean value observed in the trial.

B24. Priority question. Please provide the following mapped EQ-5D-3L data (from the EQ-5D-5L data from EMPEROR-Reduced):

a) **Average baseline EQ-5D-3L for both arms in the trial, together with respective number of observations, and statistical significance for the difference in utility at baseline across arms.**

   The mean baseline EQ-5D-3L scores in the empagliflozin and placebo arms (with standard deviations) are summarised below. The difference in means was [redacted] which was not statistically significantly different from 0 (p-value=0.496) (Table 38).

Table 38. Baseline EQ-5D-3L scores in EMPEROR-Reduced

	Empagliflozin 10 mg		Placebo		Empagliflozin 10 mg
	N	Mean (SD)	N	Mean (SD)	
Baseline EQ-5D-3L	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

b) Change from baseline in EQ-5D-3L, for both arms in the trial, together with respective number of observations, and statistical significance of changes, for all available time points (see table below for an example)

Changes from baseline EQ-5D-3L at each follow-up visit are summarised in Table 39

Table 39 Change from baseline in EQ-5D-3L score at each follow-up visit

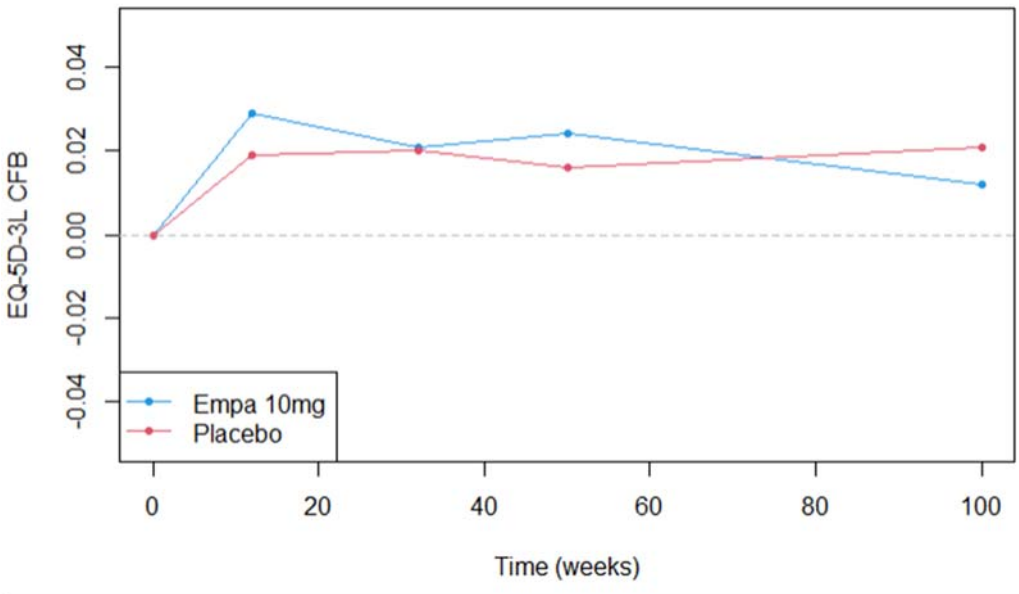
Change from Baseline	Empagliflozin 10 mg		Placebo		Difference (SE; p-value)
	N	Mean (SD)	N	Mean (SD)	
Week 12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 32	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 52	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 100	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 148[a]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

a Not included in analyses due to low patient counts.

c) A plot with the changes from baseline as requested in the previous question.

Figure 12 shows the change in utility values over time, as requested.

Figure 12. Change in utility values over time in EMPEROR-Reduced



B25. Priority question. Please run a scenario analysis where acute renal failure and hepatic injury events do not have an associated disutility (i.e. remove these disutilities) in the model, but instead, use the disutility value of 0.0762 associated with the composite renal outcome that is provided in TA679, company submission, Table 39, page 115.

We have implemented the requested scenario changing the disutility values for acute renal failure and hepatic injury events from [redacted] and -[redacted] as per Table 39, page 115 in TA679 company's submission to the requested value of -0.0762. We also changed the associated SE for both types of AEs with that reported in the same table (i.e. 0.0141). Overall, there was no impact on cost-effectiveness as the ICER changed from [redacted] to [redacted]. Therefore, this is only a minor consideration for the committee.

B26. Please combine the scenario requested in B25 with the following analysis: [redacted], please conduct a scenario analysis where the disutility values used for genital mycotic infection, volume depletion, urinary tract infection and hypoglycaemia are those provided in TA679, company submission, Table 40, page 120. The disutility value associated with bone fractures and hypertension do not need changing.

BI have implemented the requested scenario using the changes incorporated as part of B25 while also changing the disutilities associated with genital mycotic infection, volume depletion, urinary tract infection and hypoglycaemia based on those provided in TA679. The means (SE) of the AE disutilites in the original base-case and in B26 are shown in Table 40.

Table 40: Disutilities associated with AEs in the original base-case and in the scenario implemented as part of Question B26.

	Original base-case – mean (SE)	B26 – mean (SE)
Urinary Tract Infection	-0.025 (0.027)	-0.003 (0.001)
Genital Mycotic Infection	[redacted]	-0.003 (0.001)
Acute renal failure	[redacted]	-0.076 (0.014)
Hepatic injury	[redacted]	-0.076 (0.014)
Volume depletion	[redacted]	-0.051 (0.012)
Hypotension	-0.025 (0.000)	-0.025 (0.000)
Hypoglycemic event	[redacted]	-0.014 (0.001)
Bone fracture	[redacted]	-0.165 (0.037)

Abbreviations: se, standard error.

Overall, there was no impact on cost-effectiveness with the ICER changing from [redacted] to [redacted].

B27. Please explain why the Life-Years gained in the model is lower than the QALYs gained.

The life-years and QALYs per arm estimated by the economic model for each KCCQ-CSS health state are shown in Table 41. Within each health state, the total number of QALYs remains lower than the total number of life-years. Also, the incremental life-years are higher than the incremental QALYs across all health states and hence the model retains face validity.

Table 41: Overall life-years and QALYs estimated for each arm by the economic model per KCCQ-CSS health state.

	Empagliflozin + SoC	SoC	Incremental
Life-Years			
KCCQ-CSS Q1	[redacted]	[redacted]	[redacted]
KCCQ-CSS Q2	[redacted]	[redacted]	[redacted]
KCCQ-CSS Q3	[redacted]	[redacted]	[redacted]
KCCQ-CSS Q4	[redacted]	[redacted]	[redacted]
Sum	[redacted]	[redacted]	[redacted]
QALYs			
KCCQ-CSS Q1	[redacted]	[redacted]	[redacted]
KCCQ-CSS Q2	[redacted]	[redacted]	[redacted]

KCCQ-CSS Q3	[redacted]	[redacted]	[redacted]
KCCQ-CSS Q4	[redacted]	[redacted]	[redacted]
Sum	[redacted]	[redacted]	[redacted]

Abbreviations: Q, Quartile; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire clinical summary score; QALYs, quality-adjusted life-years.

However, as the ERG pointed out, the overall sum of QALYs across all health states ([redacted]) is higher than the overall sum of life-years ([redacted]). This is attributed to the fact that patients do not gain life-years in all health states. In fact, patients gain life-years in Q2 and Q4 and lose life-years in Q1 and Q3. Also, different utility values are assigned to the various KCCQ-CSS health states with higher quartiles assigned higher utility values. Hence, the final results arise as the utility-weighted sum of the life-years across all KCCQ-CSS health states. Given that patients, generally, gain life in health states that are assigned higher utility values and lose life-years in health states that are assigned lower utility values, it remains possible for the model to retain face validity and result in higher predicted overall life-years than QALYs.

To further contextualise this phenomenon, a simple fictitious example is illustrated in Table 42. In this example, patients lose on average one life year in a health state that is associated with a low utility value (0.4) and gain two life-years in a health state that is associated with a high utility value (0.8). As a result, the overall gain in QALYs is higher than the overall gain in life-years.

Table 42. An illustrative example of health states for which the overall QALYs are higher than the overall life-years

	Loss	Gain	Overall
Life-years	1	2	+1
QALYs	1*0.4=0.4	2*0.8=1.6	+1.2

Abbreviations: QALYs, quality-adjusted life-years.

Costs

B28. Priority question. The ERG disagrees with the use of the chosen estimates from the Alva paper to estimate the cost of CV deaths in the model for fatal myocardial infarction, fatal ischaemic heart disease, and fatal stroke. The company used the regression analysis presented in the paper which estimated the added inpatient costs for T2D complications. Therefore, please use the following costs (and update to the current cost year where needed) from Table 3 in the Alva paper: fatal myocardial infarction £1,521, fatal ischaemic heart disease £3,766, and fatal stroke £3,954.

If, alternatively, the company wishes to use the results from the regression analysis in Alva, please do so by weighting the CV-death costs (in patients with T2D) by the proportion of patients with T2D estimated in EMPEROR-2 with the CV-death costs mentioned in the previous paragraph (in patients without T2D) by the proportion of patients without T2D.

We have implemented the requested scenario using the suggested cost estimates for fatal myocardial infarction, fatal ischaemic heart disease, and fatal stroke. Overall, the ICER changed from [redacted] to [redacted] without any impact on cost-effectiveness conclusions.

B29. Priority question. The CV-death cost estimates from Alva capture the annual costs associated with the events leading to death. Given the company's model uses 1 month-long cycles, please conduct an adjustment in the cost estimates applied in every cycle to accurately reflect the discounting factor.

We acknowledge the fact that the Alva et al. study (63) mentions annual costs associated with events, when referring to all events for which costs were estimated by the authors, both fatal and non-fatal. However, when populating the model, the assumption has been that the costs of fatal events would be incurred close or just before the fatal event. Hence, the company believes that the discounting applied in the submitted model is correct. Further, as shown in the response to question B30 below, a reduction in the CV-death cost of up to 60% only leads to a marginal increase in ICER, from [redacted] to [redacted]. Hence, we believe that a differential

discounting applied to the CV-death cost will not have a material effect on the ICER, and thus is a minor consideration for the committee

B30. Priority question. The company assumed a [redacted]

Therefore, please conduct a scenario analysis (or consider changing your base-case) to reflect:

1. **The weighted costs of CV-death by the proportion of events leading to CV deaths observed in EMPEROR-Reduced.**
2. **The cost of sudden cardiac death as the highest contributor to the costs associated with CV deaths in the model.**

   To address the ERG's concern with regards to the weights of events leading to CV deaths in the model, we have conducted two analyses, both of which include the cost of sudden cardiac death as the highest contributor to the costs of CV deaths, as well as HF as the second highest contributor, according to Table 11.1.2.4.2: 1 in the EMPEROR Reduced clinical study report(11). In the first analysis, we conservatively assumed the cost of sudden cardiac death to be zero and noted that the ICER increased by [redacted], from [redacted] to [redacted], without any impact on the cost-effectiveness conclusions. In the second analysis, we used a unit cost of £1,632 for all sudden cardiac deaths in the model corresponding to the TOTAL HRG - National Schedule of NHS Costs (Year : 2019-20) weighted average of EB05A:C, and noted the ICER to increase by [redacted], from [redacted] to [redacted]. In both analyses, the cost of CV-death was calculated as a weighted average of: sudden cardiac death, HF, other CV causes, stroke, acute MI, using the weights reported in Table 11.1.2.4.2: 1 of EMPEROR-Reduced clinical study report, rebalanced to exclude "undetermined" and "CV procedures" for which a unit cost could not be determined, or the incidence was very low, respectively. Equally, in both analyses, the unit cost of ischaemic heart disease from Alva et al. (63) study

was assigned to “other CV causes” observed in EMPEROR-Reduced, and a unit cost of £2,061 for fatal HF was applied, using the TOTAL HRG - National Schedule of NHS Costs (Year : 2019-20) EB03A:E.

The conclusion from these scenario analyses is that the cost of fatal AE have limited impact on the ICER, and therefore is a minor consideration for the committee.

B31. Priority question. Please conduct a scenario analysis with the resource use detailed in Table A and conduct a separate scenario analysis with the outpatient/inpatient visits detailed in Table B to reflect the ERG’s clinical expert opinion.

Table A

Resource	Annual visits (company’s base-case)	ERG’s clinical expert (per year)
GP visit	23.14	5 GP visits a year and 18 nurse visits a year
Cardiologist visit	0.0504	0.2
A&E referral	0.0096	0.02

Table B

Adverse event type	% treated as Outpatient visit	% treated as Inpatient visit
Urinary tract infection	90%	10%
Genital mycotic infection	100%	0%
Acute renal failure	50%	50%
Hepatic injury	70%	30%
Volume depletion	95%	5%
Hypotension	100%	0%

Hypoglycaemic event	**90%**	**10%**
Bone fracture	**0%**	**100%**

Scenario A: The option to select alternative disease management inputs has been built into the model. The cost of the nurse visit has been sourced from Table 10.2 in the Unit Costs of Health and Social Care 2020, as £38 per hour (64). Since the cost per patient contact has not been specified, two analysis were conducted assuming the nurse visit lasts:

i) 1 hour, or
ii) 15 minutes.

Results of the two analyses are shown in Table 43. In both cases, the incremental costs associated with empagliflozin + SoC are slightly reduced compared to the base-case due to the lower cost of the nurse visit compared to the GP visit, leading to a reduction in the ICER. The general conclusions of the cost-effectiveness analysis, which suggests that empagliflozin is a cost-effective add-on to SoC in patients with HFrEF, remain unaffected.

Table 43. Deterministic cost-effectiveness of empagliflozin as an add-on to standard care, with the resource use for disease management as shown in Table A

Technology	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
Assuming 1h duration of the nurse visit							
SoC	£16,881	[redacted]	[redacted]	-	-	-	-
Empagliflozin + SoC	£17,805	[redacted]	[redacted]	£925	[redacted]	[redacted]	[redacted]
Assuming 15 min duration of the nurse visit							
SoC	£14,784	[redacted]	[redacted]				
Empagliflozin + SoC	£13,954	[redacted]	[redacted]	£830	[redacted]	[redacted]	[redacted]

Abbreviations: LYG, life-years gained; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; SoC, standard of care.

Scenario B: When the setting the management of AEs to reflect the ERG's preferred inputs shown in Table B, the ICER for empagliflozin + SoC relative to SoC alone increases to [redacted] (Table 44). The increase in ICER is driven by the increase in the cost of treating urinary tract infections and volume depletion, the AEs which occurred with higher frequency in the empagliflozin compared to the placebo arm of the EMPEROR-Reduced. The cost increase is marginal, and the conclusions of the cost-effectiveness analysis remain unchanged.

Table 44. Deterministic cost-effectiveness of empagliflozin as an add-on to standard care assuming the resource use for AE management as in Table B

Technology	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
Base-case results in the original submission							
SoC	£16,887	[redacted]	[redacted]	-	-	-	-
Empagliflozin + SoC	£17,950	[redacted]	[redacted]	£1,063	[redacted]	[redacted]	[redacted]
Updated base-case in response to B15 and B19							
SoC	£16,911	[redacted]	[redacted]				
Empagliflozin + SoC	£17,837	[redacted]	[redacted]	£926	[redacted]	[redacted]	[redacted]
Scenario assuming RU for AE management as in Table B							
SoC	[redacted]	[redacted]	[redacted]				
Empagliflozin + SoC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: AE, adverse event; ICER, incremental cost-effectiveness ratio; LYG, life-years gained; QALY, quality-adjusted life year; RU, resource use; SoC, standard of care.

Section C: Textual clarification and additional points

C1. In the company submission, Table 44, the rates of AEs for empagliflozin and SoC seem to be transposed when compared with the model estimates. Please confirm which source (Table 44 or the estimates in F174:M175 of the "Clinical inputs" tab) is correct.

[redacted] in **Table *45***.

Table 45.Rates of AEs in the modelled cohort

	Rate per 1000 patient years in the EMPEROR-Reduced trial	
	Empagliflozin + SoC	**Placebo + SoC**
Urinary tract infection	[redacted]	[redacted]
Genital mycotic infection	[redacted]	[redacted]
Acute renal failure	[redacted]	[redacted]
Hepatic injury	[redacted]	[redacted]
Volume depletion	[redacted]	[redacted]
Hypotension	[redacted]	[redacted]
Hypoglycaemic event*	[redacted]	[redacted]
Bone fracture	[redacted]	[redacted]

Abbreviation: SoC, standard of care.
* Defined as an event with a plasma glucose value of ≤70 mg/dL or where assistance was required.

C2. In the company submission, Table 45, a coefficient is listed for hypotension, but in the model the coefficient is [redacted](cell L44, tab "Risk Equations - Active". Please confirm which source (the company submission or the model estimate) is correct?

The coefficient listed in the model for hypotension is accurate and the one in Table 45 needs to be updated to 0.

C3. In the company submission, Table 48, the daily and monthly cost for furosemide does not match the model. Please confirm which source (the company submission or the model estimates) is correct?

For most patients with HFrEF, furosemide is prescribed in the community setting to treat breathlessness and oedema, as part of standard care (65). The initial daily dose is 20-40 mg, with the usual dose ranging from 40 to 240 mg (65). Since doses higher than 50 mg are administered by intravenous infusion only (66), we assumed that the dose taken orally in the community setting is on average likely to be no higher than 50 mg per day. Therefore, the daily dose of 50 mg as specified in the model is believed by the company to be accurate.

C4. In Table 23 of Document B, HHF is reported as a RR rather than a HR. Please can you check this?

Yes, the Lin-Wei-Yang-Ying model used to estimate the rate of HHF produces rate ratios (RR).

List of Appendices:

Appendix 1. Risk equations_mortality and HHF

Appendix 2. Statistical Analysis Report_Revised 30 July 2021

Appendix 3. Updated Cost-Effectiveness in response to clarification questions (30 July 2021)

Appendix 4: Empagliflozin Summary of Product Characteristics (July 2021)

References

1. IQVIA. LPD, IQVIA Ltd, incorporating data derived from THIN, A Cegedim Database, May 2021.

2. NICE. Technology appraisal guidance [TA398]: Lumacaftor–ivacaftor for treating cystic fibrosis homozygous for the F508del mutation 2016 [Available from: https://www.nice.org.uk/guidance/ta398/chapter/committee-discussion#cost-effectiveness-2.

3. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. Process and methods [PMG9]. 2013 [Available from: https://www.nice.org.uk/process/pmg9/chapter/the-reference-case.

4. National Institute for Health and Care Excellence. Final appraisal document: Dapagliflozin for treating chronic heart failure with reduced ejection fraction. 2020.

5. Compendium EM. Summary of Product Characteristics [dapagliflozin] 2021 [Available from: https://www.medicines.org.uk/emc/product/7607/smpc.

6. Summary of Product Characteristics. Jardiance 2020 [Available from: https://www.ema.europa.eu/en/documents/product-information/jardiance-epar-product-information_en.pdf.

7. Packer M, Anker SD, Butler J, Filippatos G, Pocock SJ, Carson P, et al. Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure. New England Journal of Medicine. 2020;383(15):1413-24.

8. Zannad F, Ferreira JP, Pocock SJ, Anker SD, Butler J, Filippatos G, et al. SGLT2 inhibitors in patients with heart failure with reduced ejection fraction: a meta-analysis of the EMPEROR-Reduced and DAPA-HF trials. Lancet. 2020;396(10254):819-29.

9. McMurray JJV, Solomon SD, Inzucchi SE, Kober L, Kosiborod MN, Martinez FA, et al. Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction. N Engl J Med. 2019;381(21):1995-2008.

10. National Institute for Health and Care Excellence. Chronic heart failure in adults: diagnosis and management NICE guideline [NG106] 2018 [Available from: https://www.nice.org.uk/guidance/ng106.

11. Boehringer Ingelheim International GmbH. A phase III randomised, double-blind trial to evaluate efficacy and safety of once daily empagliflozin 10 mg compared to placebo, in patients with chronic Heart Failure with reduced Ejection Fraction (HFrEF). Clinical Trial Report c28576542-012020. 2020.

12. Zannad F, Ferreira JP, Pocock SJ, Zeller C, Anker SD, Butler J, et al. Cardiac and Kidney Benefits of Empagliflozin in Heart Failure Across the Spectrum of Kidney Function: Insights From EMPEROR-Reduced. Circulation. 2021;143(4):310-21.

13. Boehringer Ingelheim. Incidence prevalence and resource utilisation of heart failure in England: a linked descriptive analysis of English primary and secondary care – PULSE study. [Data on File]. 2021.

14. Public Health England. Local action on health inequalities Understanding and reducing ethnic inequalities in health. . 2018.

15. Gov.uk. Population of England and Wales 2020 [Available from: https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/population-of-england-and-wales/latest.

16. Mathur R, Hull SA, Badrick E, Robson J. Cardiovascular multimorbidity: the effect of ethnicity on prevalence and risk factor management. Br J Gen Pract. 2011;61(586):e262-70.

17. Lam CSP, Ferreira JP, Pfarr E, Sim D, Tsutsui H, Anker SD, et al. Regional and ethnic influences on the response to empagliflozin in patients with heart failure and a reduced ejection fraction: the EMPEROR-Reduced trial. Eur Heart J. 2021.

18. Butler J, Anker SD, Filippatos G, Khan MS, Ferreira JP, Pocock SJ, et al. Empagliflozin and health-related quality of life outcomes in patients with heart failure with reduced ejection fraction: the EMPEROR-Reduced trial. Eur Heart J. 2021.

19. Singh JSS, Mordi IR, Vickneson K, Fathi A, Donnan PT, Mohan M, et al. Dapagliflozin Versus Placebo on Left Ventricular Remodeling in Patients With Diabetes and Heart Failure: The REFORM Trial. Diabetes Care. 2020;43(6):1356-9.

20. Kato ET, Silverman MG, Mosenzon O, Zelniker TA, Cahn A, Furtado RHM, et al. Effect of Dapagliflozin on Heart Failure and Mortality in Type 2 Diabetes Mellitus. Circulation. 2019;139(22):2528-36.

21. Nassif ME, Windsor SL, Tang F, Khariton Y, Husain M, Inzucchi SE, et al. Dapagliflozin Effects on Biomarkers, Symptoms, and Functional Status in Patients With Heart Failure With Reduced Ejection Fraction. Circulation. 2019;140(18):1463-76.

22. Armstrong PW, Pieske B, Anstrom KJ, Ezekowitz J, Hernandez AF, Butler J, et al. Vericiguat in Patients with Heart Failure and Reduced Ejection Fraction. N Engl J Med. 2020;382(20):1883-93.

23. Carbone S, Billingsley HE, Canada JM, Bressi E, Rotelli B, Kadariya D, et al. The Effects of Canagliflozin compared to Sitagliptin on Cardiorespiratory Fitness in Type 2 Diabetes Mellitus and Heart Failure with Reduced Ejection Fraction: Results of the CANA-HF Study. Diabetes Metab Res Rev. 2020; , http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L631838667.

24. Cosentino F, Cannon CP, Cherney DZI, Masiukiewicz U, Pratley R, Jack SD, et al. Efficacy of Ertugliflozin on Heart Failure-Related Events in Patients with Type 2 Diabetes Mellitus and Established Atherosclerotic Cardiovascular Disease: Results of the VERTIS CV Trial. Circulation. 2020; ,.

25. Edelmann F, Jaarsma T, Comin-Colet J, Schorr J, Ecochard L, Hussain RI, et al. Rationale and Study Design of OUTSTEP-HF: A Randomized Controlled Study to Assess the Effect of Sacubitril/Valsartan and Enalapril on Physical Activity measured by Accelerometry in Patients with Heart Failure with Reduced Ejection Fraction. Eur J Heart Fail. 2020;22(9):1724-33.

26. Gao Y, Xing C, Hao W, Zhao H, Wang L, Luan B, et al. The Impact of Sacrubitril/Valsartan on Clinical Treatment and hs-cTnT and NT-ProBNP Serum

Levels and the Left Ventricular Function in Patients with Chronic Heart Failure. Int Heart J. 2020;61(1):1-6.

27. Jensen J, Omar M, Kistorp C, Poulsen MK, Tuxen C, Gustafsson I, et al. Twelve weeks of treatment with empagliflozin in patients with heart failure and reduced ejection fraction: A double-blinded, randomized, and placebo-controlled trial. Am Heart J. 2020;228:47-56.

28. Khandwalla RM, Grant D, Birkeland K, Heywood JT, McCague K, Fombu E, et al. Sacubitril/valsartan versus enalapril in patients with heart failure with reduced ejection fraction: findings from awake-hf, a randomized, clinical trial using a wearable biosensor. J Am Coll Cardiol. 2019;73(9 Supplement 1):848.

29. Mordi NA, Mordi IR, Singh JS, McCrimmon RJ, Struthers AD, Lang CC. Renal and cardiovascular effects of SGLT2 inhibition in combination with loop diuretics in patients with Type 2 diabetes and chronic heart failure: The RECEDE-CHF Trial. Circulation. 2020;142:1713–24.

30. Mullasari A. Efficacy and Safety of Ivabradine Once-Daily Prolonged-Release versus Twice-Daily Immediate-Release Formulation in Patients with Stable Chronic Heart Failure with Systolic Dysfunction: A Randomized, Double-Blind, Phase 3 Non-Inferiority (PROFICIENT) Study. Cardiol Ther. 2020; 9(2):505-21.

31. Panagov ZG, Safarova ZG, Sherieva AY, Gudieva IR, Tsukkiev ZK. The effectiveness of the drug uperio and angiotensin-converting enzyme inhibitors during dilated cardiomyopathy complicated by chronic heart failure. Int J Pharm Res. 2020;12(2):411-5.

32. Tanaka A, Hisauchi I, Taguchi I, Sezai A, Toyoda S, Tomiyama H, et al. Effects of canagliflozin in patients with type 2 diabetes and chronic heart failure: a randomized trial (CANDLE). ESC heart Fail. 2020;7(4):1585-94.

33. Desai AS, Solomon SD, Shah AM, Claggett BL, Fang JC, Izzo J, et al. Effect of sacubitril-valsartan vs enalapril on aortic stiffness in patients with heart failure and reduced ejection fraction: A randomized clinical trial. JAMA. 2019;322(11):1077-84.

34. Felker GM, McMurray J, Cleland J, O'Connor C, Teerlink J, Voors A, et al. Hemodynamic effects of nitroxyl donor bms-986231 compared to nitroglycerin or placebo in patients with chronic heart failure-a randomized double-blind crossover trial. J Am Coll Cardiol. 2020;75(11):850.

35. Kang DH, Park SJ, Shin SH, Hong GR, Lee S, Kim MS, et al. Angiotensin Receptor Neprilysin Inhibitor for Functional Mitral Regurgitation: PRIME Study. Circulation. 2019;139(11):1354-65.

36. Nassif ME, Windsor S, Tang F, Khariton Y, Husain M, Inzucchi S, et al. Dapagliflozin effects on biomarkers, symptoms, and functional status in patients with heart failure with reduced ejection fraction. Circulation. 2019;140(18).

37. Tsutsui H, Momomura SI, Yamashina A, Shimokawa H, Kihara Y, Saito Y, et al. Efficacy and safety of ivabradine in Japanese patients with chronic heart failure ? J-SHIFT study ? Circulation. 2019;83(10):2049-60.

38. Villacorta AS, Villacorta H, Caldas JA, Precht BC, Porto PB, Rodrigues LU, et al. Effects of Heart Rate Reduction With Either Pyridostigmine or Ivabradine in Patients With Heart Failure: A Randomized, Double-Blind Study. J Cardiovasc Pharmacol Ther. 2019;24(2):139-45.

39. Raja DC, Kapoor A, Sinha A, Kashyap S, Khanna R, Kumar S, et al. Heart rate manipulation in dilated cardiomyopathy: Assessing the role of Ivabradine. Indian Heart J. 2018;70(2):246-51.

40. NCT02788656. Pulmonary Artery Pressure Reduction With ENTresto (Sacubitril/Valsartan) 2016 [Available from: https://clinicaltrials.gov/show/NCT02788656.

41. Sallam M, Al-Saadi T, Alshekaili L, Al-Zakwani I. Impact of ivabradine on health-related quality of life of patients with ischaemic chronic heart failure. Curr Vasc Pharmacol. 2016;14(5):481-6.

42. Senni M, McMurray JJV, Wachter R, McIntyre HF, Reyes A, Majercak I, et al. Initiating sacubitril/valsartan (LCZ696) in heart failure: results of TITRATION, a double-blind, randomized comparison of two uptitration regimens. Eur J Heart Fail. 2016;18(9):1193-202.

43. Teerlink JR, Felker GM, McMurray JJV, Solomon SD, Adams KF, Cleland JGF, et al. Chronic Oral Study of Myosin Activation to Increase Contractility in Heart Failure (COSMIC-HF): a phase 2, pharmacokinetic, randomised, placebo-controlled trial. Lancet. 2016;388(10062):2895-903.

44. Tsutsui H, Momomura S, Yamashina A, Ogawa H, Shimokawa H, Kihara Y, et al. Heart rate control with if inhibitor, ivabradine, in Japanese patients with chronic heart failure–A randomized, double-blind, placebo-controlled phase II study–. Circulation. 2016;80(3):668-76.

45. Abdel-Salam Z, Rayan M, Saleh A, Abdel-Barr MG, Hussain M, Nammas W. If current inhibitor ivabradine in patients with idiopathic dilated cardiomyopathy: Impact on the exercise tolerance and quality of life. Cardiol J. 2015;22(2):227-32.

46. Amosova KM, Katsytadze II, Rudenko YU, Bezrodniy A. Early effects of Ivabradin-b-blocker combination compared to b-blocker uptitration on LV and pulsatile arterial functions in postinfarction patients with mild CHF and EF lowering. Eur J Heart Fail. 2015;17:341-2.

47. Gheorghiade M, Greene SJ, Butler J, Filippatos G, Lam CSP, Maggioni AP, et al. Effect of vericiguat, a soluble guanylate cyclase stimulator, on natriuretic peptide levels in patients withworsening chronic heart failure and reduced ejection fraction the socrates-reduced randomized trial. JAMA. 2015;314(21):2251-62.

48. Greenberg BH, Chou W, Saikali KG, Escandón R, Lee JH, Chen MM, et al. Safety and tolerability of omecamtiv mecarbil during exercise in patients with ischemic cardiomyopathy and angina. JACC: Heart Fail. 2015;3(1):22-9.

49. Lopatin YM, Temirsultanova TH. Tight heart rate control and pulsatile hemodynamics in patients with heart failure of ischemic etiology. Eur Heart J. 2015;36:665.

50. Ordu S, Yildiz BS, Alihanoglu YI, Ozsoy A, Tosun M, Evrengul H, et al. Effects of ivabradine therapy on heart failure biomarkers. Cardiol J. 2015;22(5):501-9.

51. Chaudhari MP, Chakraborty P, Pathak SP, Isser HS, Bansal S. Effects of ivabradine on left ventricular function in patients with ischemic heart failure. Indian Heart J. 2014;66:S83?S4.

52. McMurray JJV, Packer M, Desai AS, Gong J, Lefkowitz MP, Rizkala AR, et al. Angiotensin-neprilysin inhibition versus enalapril in heart failure. N Engl J Med. 2014;371(11):993-1004.

53. Moiseev V, Potapenko A, Abdulasisov O, Dyatchuk L, Meray I, Kiyakbaev G, et al. Efficacy of Ivabradine in patients with chronic heart failure after myocardial infarction. Eur J Heart Fail , Suppl. 2011;10:S234.

54. Volterrani M, Cice G, Caminiti G, Vitale C, D'Isa S, Filardi PP, et al. Effect of Carvedilol, Ivabradine or their combination on exercise capacity in patients with Heart Failure (the CARVIVA HF trial). Int J Cardiol. 2011;151(2):218-24.

55. Sarullo FM, Fazio G, Puccio D, Fasullo S, Paterna S, Novo S, et al. Impact of "off-label" use of ivabradine on exercise capacity, gas exchange, functional class, quality of life, and neurohormonal modulation in patients with ischemic chronic heart failure. J Cardiovasc Pharmacol Ther. 2010;15(4):349-55.

56. Swedberg K, Komajda M, Böhm M, Borer JS, Ford I, Dubost-Brama A, et al. Ivabradine and outcomes in chronic heart failure (SHIFT): A randomised placebo-controlled study. Lancet. 2010;376(9744):875-85.

57. Fox K, Ford I, Steg PG, Tendera M, Ferrari R, Investigators B. Ivabradine for patients with stable coronary artery disease and left-ventricular systolic dysfunction (BEAUTIFUL): a randomised, double-blind, placebo-controlled trial. Lancet. 2008;372(9641):807-16.

58. Taylor AL, Ziesche S, Yancy C, Carson P, Jr RDA, Ferdinand K, et al. Combination of isosorbide dinitrate and hydralazine in blacks with heart failure. N Engl J Med. 2004;351(20):2049-141.

59. Cohn JN, Johnson G, Ziesche S, Cobb F, Francis G, Tristani F, et al. A comparison of enalapril with hydralazine-isosorbide dinitrate in the treatment of chronic congestive heart failure. N Engl J Med. 1991;325(5):303-10.

60. Lin M, Chiang HT, Chen CY, Chiang BN. Comparison of enalapril and conventional vasodilator therapy in patients with chronic congestive heart failure. J Formos Med Assoc/ Taiwan yi zhi. 1991;90(5):452-9.

61. Schofield PM, Brooks NH, Lawrence GP, Testa HJ, Ward C. Which vasodilator drug in patients with chronic heart failure? A randomised comparison of captopril and hydralazine. Br J Clin Pharmacol. 1991;31(1):25-32.

62. Cohn JN, Archibald DG, Ziesche S. Effect of vasodilator therapy on mortality in chronic congestive heart failure: Results of a Veterans Administration Cooperative Study. N Engl J Med. 1986;314(24):1547-52.

63. Alva ML, Gray A, Mihaylova B, Leal J, Holman RR. The impact of diabetes-related complications on healthcare costs: new results from the UKPDS (UKPDS 84). Diabetic Medicine. 2015;32(4):459-66.

64. Unit Costs of Health & Social Care 2020 [Internet]. Unit Costs of Health and Social Care . PSSRU, University of Kent. 2020. Available from: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2020/.

65. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart

failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. European Heart Journal. 2016;37(27):2129-200.

66. NICE. Indication and doses: National Institute for Health and Care Excellence, British National Formulary; [Available from: https://bnf.nice.org.uk/drug/furosemide.html#indicationsAndDoses.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Empagliflozin for treating chronic heart failure with reduced ejection fraction [ID3826]

Additional Clarification questions

August 2021

File name	Version	Contains confidential information	Date
ID3826 Empagliflozin clarification questions_12_08_21_FINAL	1.0	Yes	12th August 2021

Notes for company

Highlighting in the template

Square brackets and grey highlighting are used in this template to indicate text that should be replaced with your own text or deleted. These are set up as form fields, so to replace the prompt text in [grey highlighting] with your own text, click anywhere within the highlighted text and type. Your text will overwrite the highlighted section.

To delete grey highlighted text, click anywhere within the text and press DELETE.

Following submission of a response to clarification questions on 5th August 2021, NICE and the ERG asked the following questions.

1) **The "new question added by the ERG" in Section B requested that the company provided the number of observations and mean KCCQ scores for the excluded period from their analysis (i.e. post 52 weeks). Therefore, can we please ask that you add these data for the end of treatment and follow-up visits in their tables 23 and 24.**

New Question

From page 100 in the clinical study report, it seems that approximately 68% of placebo patients and 61% of empagliflozin patients were still on-treatment/being observed at 52 weeks in the EMPEROR trial. The ERG also notes that the mean time on-treatment in trial was 63 weeks, and mean follow-up time was 67 weeks. Please provide:

a) Number of observations at week 12, week 32, week 52, end of treatment visit, and follow-up visit in the KCCQ-CSS dataset used in the economic analysis to estimate the transition probabilities between the KCCQ-CSS states of the model (without imputed values);
b) The mean and respective standard deviation (SD) KCCQ-CSS data (without imputed values) at week 12, week 32, week 52, end of treatment visit, and follow-up visit in the KCCQ-CSS dataset used in the economic analysis to estimate the transition probabilities between the KCCQ-CSS states of the model;
c) The number of observations and mean (and respective SD) KCCQ-CSS data (with imputed values) for week 12, week 32 and week 52 underpinning the KCCQ data used in the analysis to estimate the transition probabilities in the model;
d) The details, data used and the results of the exploratory analysis conducted by the company, described in the Evidera appendix

(page 6) which concluded that “KCCQ-CSS health states tend to change early on after start of treatment and stabilise fairly early

As requested, the mean (SD) KCCQ-CSS scores post 52 weeks are reported in Table 1 (without imputation) and

Table ***2*** (with imputation).

The post week 52 observations were not used in deriving the transition matrices for the cost-effective model because there were low number of available observations (~600 patients in each treatment arm) and imputation was required for extended periods of time. Visits post week 52 did not have a fixed scheduled time. In the EMPEROR-Reduced, KCCQ was recorded on Week 1, 12, 32, 52, End of Treatment (EOT) and EOT + 30 days(1). In both the imputed and non-imputed analysis, measurements from day 436 (week 62) were recorded as “post week 52”. In the imputed analysis, where patients were still alive and followed on day 436 but did not have a post week 52 measurement, the last available measurement was carried forward (last observation carried forward, LOCF). This is the same imputation rule used for the estimation of the transition matrices for the cost-effective model.

Table 1. Mean and SD for KCCQ-CSS data (without imputed values) as used in the economic model.

		Placebo		Empagliflozin 10 mg	
	KCCQ-CSS Score without imputation	**N**	**Mean (SD)**	**N**	**Mean (SD)**
Used in economic model	**Baseline**	[redacted]	[redacted]	[redacted]	[redacted]
	Week 12	[redacted]	[redacted]	[redacted]	[redacted]
	Week 32	[redacted]	[redacted]	[redacted]	[redacted]
	Week 52	[redacted]	[redacted]	[redacted]	[redacted]

		Placebo		Empagliflozin 10 mg	
Not used in economic model	**Post week 52**	[redacted]	[redacted]	[redacted]	[redacted]

*Estimated from the AC-OD dataset

Table 2. Mean and SD for KCCQ-CSS data (with imputed values) as used in the economic model.

		Placebo		Empagliflozin 10 mg	
	KCCQ-CSS Score with imputation	**N**	**Mean (SD)**	**N**	**Mean (SD)**
Used in economic model	**Baseline**	[redacted]	[redacted]	[redacted]	[redacted]
	Week 12	[redacted]	[redacted]	[redacted]	[redacted]
	Week 32	[redacted]	[redacted]	[redacted]	[redacted]
	Week 52	[redacted]	[redacted]	[redacted]	[redacted]
Not used in economic model	**Post week 52**	[redacted]	[redacted]	[redacted]	[redacted]

2) It seems that the company might have a new base case, as per results in the model and (for example) the company's answer to CQ B19. Could you please ask the company to confirm what the new base case ICER is, and send updated tables and graphs with cost-effectiveness results, sensitivity (particularly PSA) and scenario analyses incorporating the revised base-case ICER.

In Question B19, the ERG requested that BI remove the half-cycle correction applied to the discounting factor on costs and benefits (i.e. in cycle 0 of the model the discounting factor should be 1 and not 0.999, etc.).

The ERG asked BI to update the base case to reflect this. We call this the "updated base case". The original base case is the ICER submitted by the company on the 30th June 2021.

Compared to the original base case, the deterministic ICER in the updated base case decreased by £87/QALY. The deterministic ICER in the updated base case was £4,717 compared to £4,804/QALY in the original base case (Table 3). These differences were marginal and indicated that amendments in how the half cycle correction was implemented had limited impact on cost effectiveness.

A probabilistic sensitivity analysis was undertaken for the updated base case. The covariance matrices for parameters informing the rate of CV-death, all-cause death, HHF and baseline utility estimates were the same as in the original base case. Similar to the original base case, all cost parameters were assigned a gamma distribution, while disutilities associated with AE's and HHF were assigned a beta distribution. Similar to the original base case, one thousand iterations were run.

The probabilistic ICER in the updated base case was broadly comparable to the original base case, indicating that changes in how half cycle correction was implemented had limited impact on cost effectiveness. Compared to the original base case, the updated base case decreased the probabilistic ICER by £87/QALY (£4,894/QALY gained vs £4,807/QALY gained, respectively) (Table 4). The results of the PSA for the updated base case are summarised in the cost-effectiveness scatterplot (Figure 1). Each point on the chart represents a single probabilistic iteration of the model. Of one thousand iterations, in the updated base case [redacted] of iterations produced ICERs that fell below a willingness-to-pay threshold of £20,000 per QALY, compared to [redacted] in the original base case (

*Figure **2***). The cost-effectiveness acceptability curve for the updated base case (Figure 3) was comparable to the base case submitted on 30th June (Figure 4). At a willingness-to-pay threshold of £4,500 per QALY, empagliflozin + SoC reaches a [redacted] probability of being cost-effective in both the updated base case and original base case. The conclusion in the original submission (30th June 2021) was that the similarity between the deterministic and the probabilistic ICERs indicated that the model is sufficiently linear. This conclusion still stands in the updated base case.

A deterministic sensitivity analysis was undertaken for the updated base case. Similar to the original base case, the most influential parameter in the updated base case was the treatment effect of empagliflozin+SoC associated with HHF. When this parameter was set to zero, the ICER was [redacted] QALY gained and [redacted] QALY gained, respectively (

Table ***5***), indicating a marginal difference of [redacted]. In the original base case, other drivers of cost effectiveness identified, including discount rates and health outcomes, and the treatment effect of empagliflozin+SoC associated with all-cause mortality. These were still important drivers of cost effectiveness in the updated base case (

Table ***5***, Figure 5, Figure 6).

Finally a scenario analysis was undertaken for the updated base case (Table 6). The results show that the ICERs for the scenario analysis for the original and updated base case were comparable.

Table 3. Base-case analysis: deterministic results for empagliflozin as an add-on to standard care (submitted on 30th June vs updated post clarification questions)

	Empagliflozin + SoC	SoC	Incremental
Updated base case following clarification questions (Q19 – amendment to half cycle correction)			

Total costs (£)	£17,837	£16,911	£926
Total LYG	5.81	5.63	0.18
Total QALYs	3.76	3.56	0.20
Cost per QALY gained	-	-	£4,717
Cost per LY gained	-	-	£5,089
Cost per HHF avoided	-	-	£5,335
Numbers needed to treat (HHF)	-	-	6.00
Numbers needed to treat (CV-death)	-	-	140.00
Company submission (submitted on 30th June 2021)			
Total costs (£)	£17,950	£16,887	£1,063
Total LYG	5.83	5.62	0.21
Total QALYs	3.78	3.55	0.22
Cost per QALY gained	-	-	£4,804
Cost per LY gained	-	-	£5,173
Cost per HHF avoided	-	-	£5,229
Numbers needed to treat (HHF)	-	-	5.00
Numbers needed to treat (CV-death)	-	-	118.00

QALY, quality-adjusted life year

Table 4. Base-case analysis: probabilistic results for empagliflozin as an add-on to standard care (submitted on 30th June vs updated post clarification questions)

	Empagliflozin + SoC	SoC	Incremental
Updated base case following clarification questions (Q19 – half-cycle correction)			
Total costs (£)	£17,719	£16,795	£923.59
Total LYG	5.75	5.57	0.18
Total QALYs	3.72	3.53	0.19
Cost per QALY gained	-	-	£4,807
Percentage of replications cost effective at £20,000/QALY, %	-	-	83%
Company submission (submitted on 30th June 2021)			
Total costs (£)	£17,876	£16,830	£1,046
Total LYG	5.76	5.56	0.20
Total QALYs	3.74	3.52	0.21
Cost per QALY gained	-	-	4,894
Percentage of replications cost effective at £20,000/QALY, %	-	-	79%

Figure 1.Updated base-case analysis (following clarification questions): cost-effectiveness scatterplot

Figure 2.Original base-case analysis (Company submission 30th June 2021): cost-effectiveness scatterplot

Figure 3. Updated base-case analysis (following clarification questions): cost-effectiveness acceptability curve

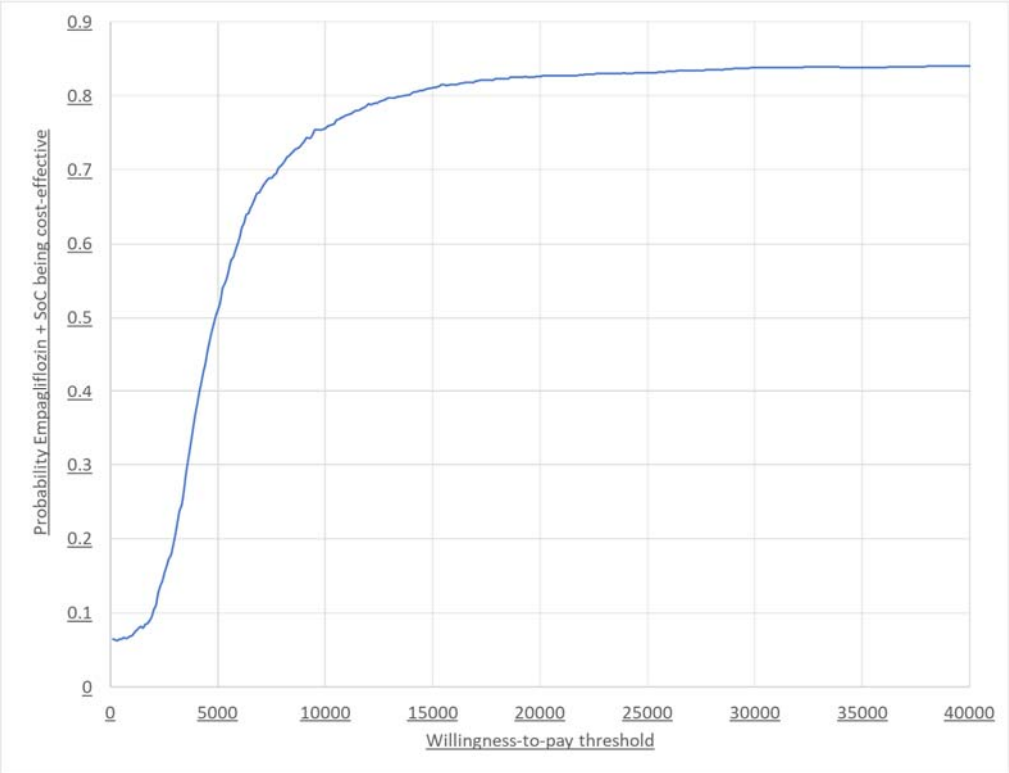


Figure 4. Original base-case analysis (Company submission 30th June 2021): cost-effectiveness acceptability curve

Table 5. Deterministic sensitivity analyses inputs and results

Scenario	Base-Case Input	Alternative Input	Description	ICER per QALY *Original base case*	ICER per QALY *Updated base case*
Base-case	-	-	-		
Clinical Inputs					
CV & all-cause death: Distribution	Weibull	Exponential	Alternative distribution	[redacted]	[redacted]
CV death: Treatment effect	[redacted]	[redacted]	No treatment effect	[redacted]	[redacted]
All-cause death: Adjust with UK lifetable?	Yes	No	No lifetable adjustment	[redacted]	[redacted]
All-cause death: Treatment effect	[redacted]	[redacted]	No treatment effect	[redacted]	[redacted]
HHF: Treatment effect	[redacted]	[redacted]	No treatment effect	[redacted]	[redacted]
Discontinuation: Distribution	Exponential	Weibull	Alternative distribution	[redacted]	[redacted]
Discontinuation: Treatment Effect	[redacted]	[redacted]	No treatment effect	[redacted]	[redacted]
Include discontinuation?	Yes	No	No discontinuation of empagliflozin	[redacted]	[redacted]
HR for empagliflozin + SoC composite renal endpoint	0.5	0.32	Lower bound of the 95% CI	[redacted]	[redacted]
		0.77	Upper bound of the 95% CI	[redacted]	[redacted]
Costs and Resource Use					
Cost of HHF	£3,072	£2,426.27	Decrease by 20%	[redacted]	[redacted]
		£3,639.40	Increase by 20%	[redacted]	[redacted]
Cost of CV death	£4,146	£3,316.80	Decrease by 20%	[redacted]	[redacted]
		£4,975.20	Increase by 20%	[redacted]	[redacted]
Unit Costs of Disease monitoring	Multiple Values	Multiple Values	Decrease by 20%	[redacted]	[redacted]
			Increase by 20%	[redacted]	[redacted]
Monthly Cost of Disease Monitoring:	£77	£62	Decrease by 20%	[redacted]	[redacted]
		£93	Increase by 20%	[redacted]	[redacted]

Scenario	Base-Case Input	Alternative Input	Description	ICER per QALY *Original base case*	ICER per QALY *Updated base case*
KCCQ-CSS 1st Quartile					
Monthly Cost of Disease Monitoring: KCCQ-CSS 2nd Quartile	£77	£62	Decrease by 20%	[redacted]	[redacted]
		£93	Increase by 20%	[redacted]	[redacted]
Monthly Cost of Disease Monitoring: KCCQ-CSS 3rd Quartile	£77	£62	Decrease by 20%	[redacted]	[redacted]
		£93	Increase by 20%	[redacted]	[redacted]
Monthly Cost of Disease Monitoring: KCCQ-CSS 4th Quartile	£77	£62	Decrease by 20%	[redacted]	[redacted]
		£93	Increase by 20%	[redacted]	[redacted]
Cost of AE management	Multiple Values	Multiple Values	Decrease by 20%	[redacted]	[redacted]
			Increase by 20%	[redacted]	[redacted]
Drug costs source	MIMS	eMIT	Alternate data source	[redacted]	[redacted]
Composite renal outcomes – annual cost	£4,862	£3,890	Decrease by 20%	[redacted]	[redacted]
		£5,834	Increase by 20%	[redacted]	[redacted]
Utilities					
Utility: KCCQ-CSS 1st Quartile	[redacted]	[redacted]	Lower 95% CI	[redacted]	[redacted]
		[redacted]	Upper 95% CI	[redacted]	[redacted]
Utility: KCCQ-CSS 2nd Quartile	[redacted]	[redacted]	Lower 95% CI	[redacted]	[redacted]
		[redacted]	Upper 95% CI	[redacted]	[redacted]
Utility: KCCQ-CSS 3rd Quartile	[redacted]	[redacted]	Lower 95% CI	[redacted]	[redacted]
		[redacted]	Upper 95% CI	[redacted]	[redacted]
Utility: KCCQ-CSS 4th Quartile	[redacted]	[redacted]	Lower 95% CI	[redacted]	[redacted]
		[redacted]	Upper 95% CI	[redacted]	[redacted]
Disutility: HHF	[redacted]	[redacted]	Lower 95% CI	[redacted]	[redacted]
		[redacted]	Upper 95% CI	[redacted]	[redacted]
Settings					
Time horizon	Lifetime	10 years	Lower range	[redacted]	[redacted]
		20 years	Upper range	[redacted]	[redacted]
Discount rate: cost	3.5%	0%	Lower range	[redacted]	[redacted]
		5%	Upper range	[redacted]	[redacted]

Scenario	Base-Case Input	Alternative Input	Description	ICER per QALY *Original base case*	ICER per QALY *Updated base case*
Discount rate: health	3.5%	0%	Lower range	[redacted]	[redacted]
		5%	Upper range	[redacted]	[redacted]

Abbreviations: AEs, adverse events; CV, cardiovascular; HR, hazard ratio; HHF, hospitalisation for heart failure; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire Clinical Summary Score; UK, United Kingdom.

Figure 5. Tornado diagram (updated base case)

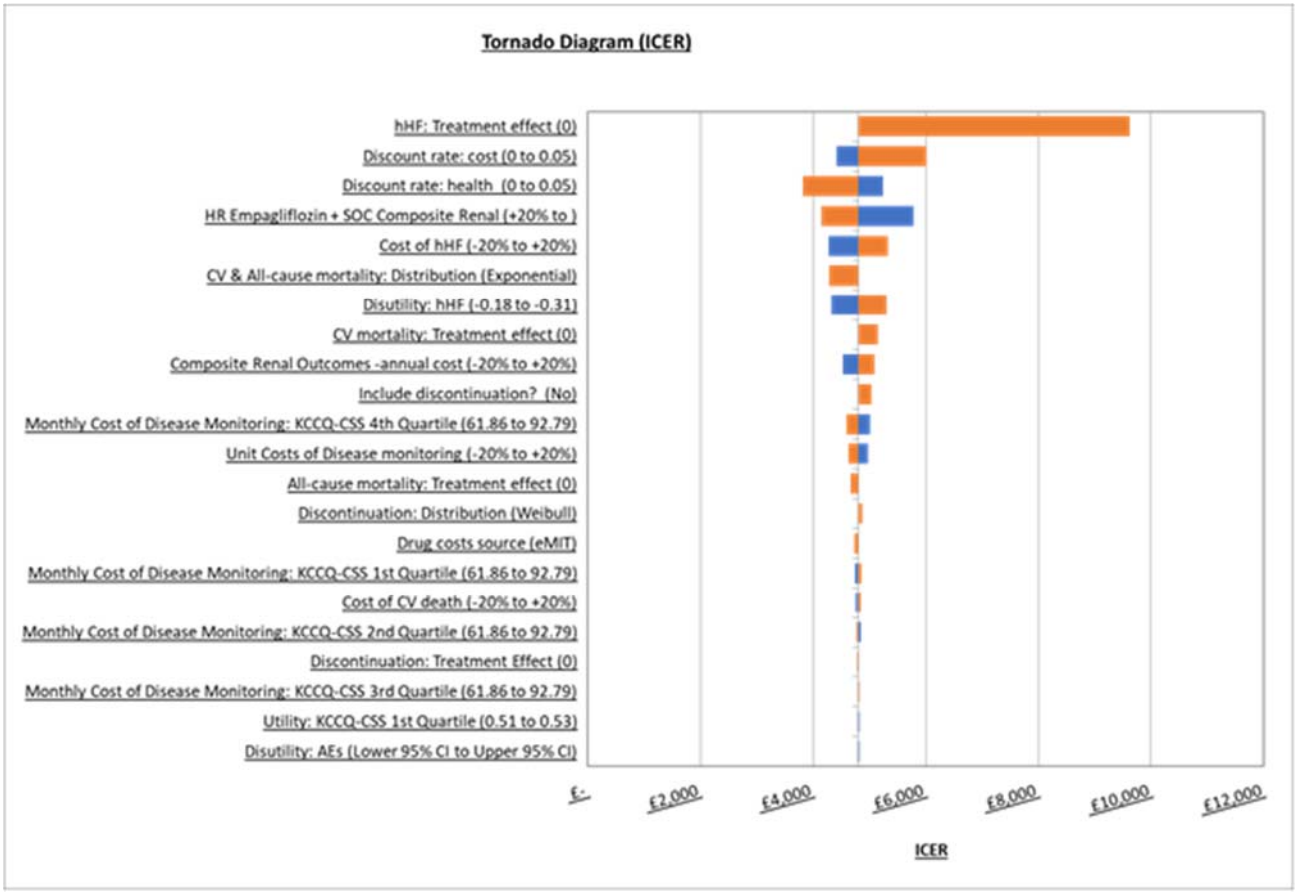
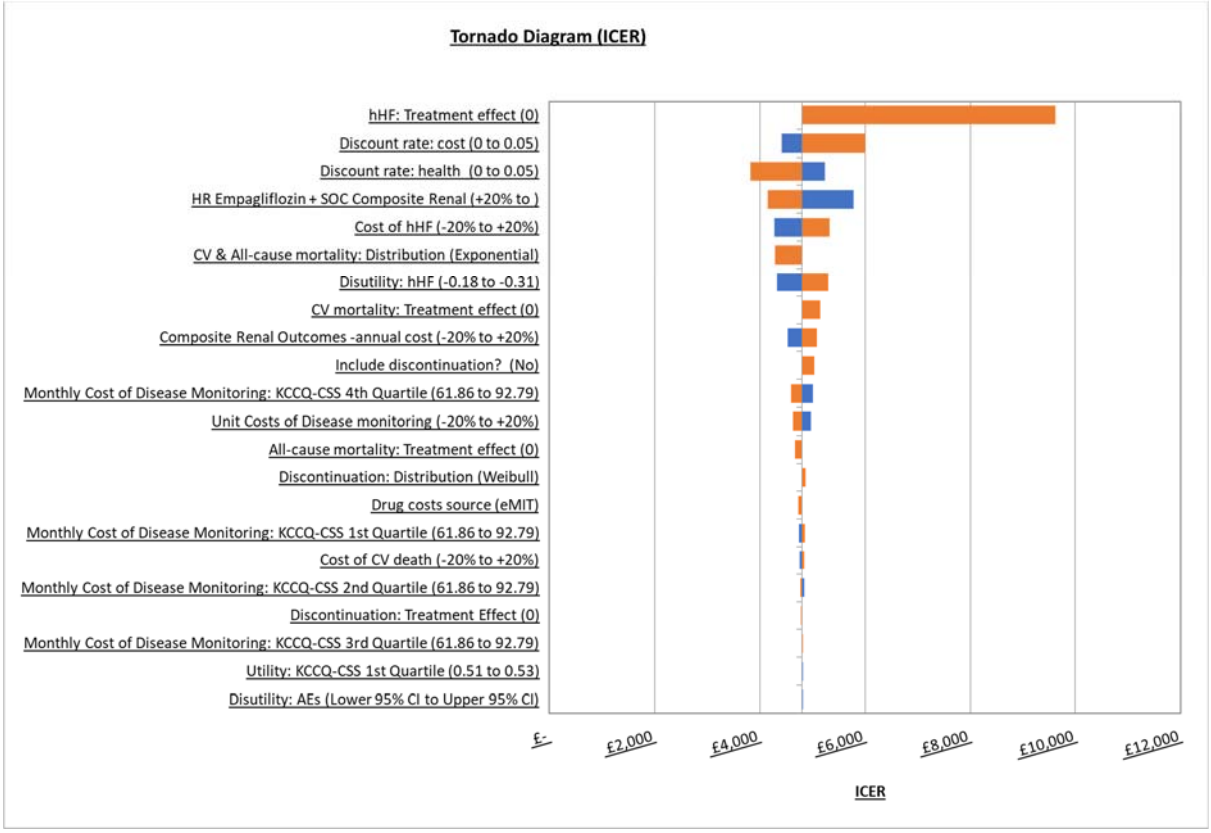


Figure 6.Tornado diagram (original base case)



Key: ICER, Incremental cost-effectiveness ratio

Table 6. Scenario Analyses: ICERs for empagliflozin as add-on to standard care compared to standard care alone

Scenario	Description	ICER (Cost in £ / QALY) Original base case	% Change Relative to original base-case ICER	ICER (Cost in £ / QALY) Updated base case	% Change Relative to updated base-case ICER
Base-case		[redacted]	-	£4,717	-
SGLT2i class effect	Use of the hazard ratios from Zannad et al (2) to inform all-cause mortality, CV mortality, and HHF with SGLT2i vs SoC based on pooled data from DAPA-HF and EMPEROR-Reduced.	[redacted]	[redacted]	[redacted]	2.52%
One Inflection Point	Use the KCCQ quartile transition matrix used for months 4 – 8 in the model base-case from month 4 to the end of the time horizon.	[redacted]	[redacted]	[redacted]	14.35%
Mortality: log normal	Extrapolate CV and all-cause mortality outcomes using a log normal distribution.	[redacted]	[redacted]	[redacted]	-26.37%
Mortality: log-logistic	Extrapolate CV and all-cause mortality outcomes using a log-logistic distribution.	[redacted]	[redacted]	[redacted]	-12.15%
Mortality: Exponential	Extrapolate CV and all-cause mortality outcomes using an exponential distribution.	[redacted]	[redacted]	[redacted]	-10.24%
Mortality: Generalised Gamma	Extrapolate CV and all-cause mortality outcomes using a generalised gamma distribution.	[redacted]	[redacted]	[redacted]	0.42%
Mortality: Gompertz	Extrapolate CV and all-cause mortality outcomes using a Gompertz distribution.	[redacted]	[redacted]	[redacted]	11.09%
Discontinuation: Weibull	Extrapolate time to discontinuation for empagliflozin using a Weibull distribution.	[redacted]	[redacted]	[redacted]	0.98%
Discontinuation: log normal	Extrapolate time to discontinuation for empagliflozin using a log normal distribution.	[redacted]	[redacted]	[redacted]	2.29%

Discontinuation: log-logistic	Extrapolate time to discontinuation for empagliflozin using a log-logistic distribution.	[redacted]	[redacted]	[redacted]	1.57%
Discontinuation: Generalised Gamma	Extrapolate time to discontinuation for empagliflozin using a generalised gamma distribution.	[redacted]	[redacted]	[redacted]	2.18%
Discontinuation: Gompertz	Extrapolate time to discontinuation for empagliflozin using a Gompertz distribution.	[redacted]	[redacted]	[redacted]	-0.55%
Utility: age-adjustment off	Use utility data as collected in the trial [redacted], without adjusting KCCQ 4 to be equal to UK general population utility.	[redacted]	[redacted]	[redacted]	-7.23%
Non-CV death costs	Assuming that non-CV deaths incur the same costs as CV deaths.	[redacted]	[redacted]	[redacted]	1.84%
No composite renal outcome costs and benefits	Excluding the costs and benefits of the composite renal outcome.	[redacted]	[redacted]	[redacted]	31.52%

Abbreviations: SoC, standard of care; SGLT2i, sodium-glucose co-transporter-2 inhibitors.
The time to discontinuation used the re-paremeterised version of the CE model – See *Table 7*

Updates to the CE model

During clarification questions, an update model was shared with the ERG and NICE. The updated model implemented all of the scenarios requested by the ERG as user selectable options. The file name was "*Appendix 3_ID3826_CEM_V2.0_AIC_Evidera_30July2021_Clarification Questions".*

A further two updates have been made to the model post submission of clarification questions. Neither of these updates impact the base case ICER. The file name is "*Appendix 3_ID3826_CEM_V2.0_AIC_Evidera_12August2021_Clarification Questions*".

1. Replaced parameter for the generalised gamma for the time to treatment discontinuation, to ensure all parameters from Appendix 1 have been transcribed into the CE model **Table *7***.
2. Amended macro so that the reporting of the % of iterations in the NE quadrant is automated (PSA!D41).

Table 7. Re-parameteristion of the time to discontination (generalised gamma)

	P1	P2	P3	Time effect	KCCQ Q2	KCCQ Q3	KCCQ Q4
Original base case	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Updated base case	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

References

1. Boehringer Ingelheim International GmbH. A phase III randomised, double-blind trial to evaluate efficacy and safety of once daily empagliflozin 10 mg compared to placebo, in patients with chronic Heart Failure with reduced Ejection Fraction (HFrEF). Clinical Trial Report c28576542-012020. 2020.

2. Zannad F, Ferreira JP, Pocock SJ, Anker SD, Butler J, Filippatos G, et al. SGLT2 inhibitors in patients with heart failure with reduced ejection fraction: a meta-analysis of the EMPEROR-Reduced and DAPA-HF trials. Lancet. 2020;396(10254):819-29.

NICE National Institute for Health and Care Excellence

Patient organisation submission

Empagliflozin for treating chronic heart failure with reduced ejection fraction [ID3826]

Thank you for agreeing to give us your organisation's views on this technology and its possible use in the NHS.

You can provide a unique perspective on conditions and their treatment that is not typically available from other sources.

To help you give your views, please use this questionnaire with our guide for patient submissions.

You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type. [Please note that declarations of interests relevant to this topic are compulsory].

Information on completing this submission

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 10 pages.

About you	
1.Your name	[redacted]

2. Name of organisation	Cardiomyopathy UK
3. Job title or position	[redacted]
4a. Brief description of the organisation (including who funds it). How many members does it have?	Cardiomyopathy UK is the national charity for people affected by the heart muscle disease, cardiomyopathy. The charity provides direct support via its services (Helpline, Peer support, resources) to over 5,000 people per year and a further 500,000 people access support online. In addition to providing support services the charity provided healthcare professional education, raises awareness of the condition, support research and advocates for improved access to treatment. Funding predominately comes from the cardiomyopathy community as well as charitable funders such as the National Lottery Community Fund and BBC's Children in Need. Around 15% of the charity's total income comes from the pharmaceutical industry.
4b. Has the organisation received any funding from the manufacturer(s) of the technology and/or comparator products in the last 12 months? [Relevant manufacturers are listed in the appraisal matrix.]	No

If so, please state the name of manufacturer, amount, and purpose of funding.	
4c. Do you have any direct or indirect links with, or funding from, the tobacco industry?	No
5. How did you gather information about the experiences of patients and carers to include in your submission?	Information has been taken from the charity's 2019 service user survey (n593) of people with cardiomyopathy. This survey covers user experience of diagnosis and treatment as well as living with cardiomyopathy and heart failure.7% of respondents have also been diagnosed with diabetes. Information has also be gathered via the charity's peer support network, helpline, online Facebook group and 2016 survey on the emotional health impact of cardiomyopathy (n455)
Living with the condition	
6. What is it like to live with the condition? What do carers experience when caring for someone with the condition?	The physical symptoms associated with heart failure are well documented and classified using the NYHA classification, however the mental health impact of the condition cannot be underestimated. The charity's 2016 survey on the mental health impact of cardiomyopathy (that included 116 respondents with heart failure) showed that 53% of respondents reported that their condition had a serious negative impact on their mental health. 60% of carers of people with the condition reported the same.

	The nature of this impact is best illustrated by the below comments from respondents: “My confidence has been shattered and I do not function as well as would normally.” “I think about dying a lot.” “It has affected me in the past for several weeks, a few months at a time.” “It affects different areas of my life at different times” “…Can’t do anything but wait and try not to worry”.
Current treatment of the condition in the NHS	
7. What do patients or carers think of current treatments and care available on the NHS?	While there is awareness among our community of the presence of heart failure medication (including newer treatments such as Entresto), there is a continued perception that even with these treatments heart failure is start of an irreversible and steady decline
8. Is there an unmet need for patients with this condition?	Medication is already available for heart failure but it is the speed at which individuals with heart failure are diagnosed or referred on for diagnosis from primary care which is the key issue. This is especially a problem for people who have developed heart failure from condition such as cardiomyopathy where patients may not fit the profile (age, lifestyle etc.) of what a typical person with heart failure “looks like”. Around one third of people with cardiomyopathy are initially diagnosed with condition such as asthma or anxiety.

Advantages of the technology	
9. What do patients or carers think are the advantages of the technology?	The key advantage of this treatment is that it is already widely used for diabetes in primary care. Expanding usage to heart failure will cement a connection between the two conditions and encourage GP's to consider heart disease in a larger population. The charity believes that this will ultimately lead to a reduction in misdiagnosis/delayed diagnosis of heart failure in primary care.
Disadvantages of the technology	
10. What do patients or carers think are the disadvantages of the technology?	N/A
Patient population	
11. Are there any groups of patients who might benefit more or less from the technology than others? If so, please describe them and explain why.	N/A

Equality	
12. Are there any potential equality issues that should be taken into account when considering this condition and the technology?	N/A
Other issues	
13. Are there any other issues that you would like the committee to consider?	N/A
Key messages	

15. In up to 5 bullet points, please summarise the key messages of your submission:

- Heart failure has a significant impact on both physical and mental health.
- Delayed and misdiagnosis of heart failure in primary care is a significant issue
- Expanding usage of this treatment to the heart failure community will ultimately lead to a reduction in delayed and misdiagnosis
-

-

Thank you for your time.

Please log in to your NICE Docs account to upload your completed submission.

.....

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

.....

NICE National Institute for Health and Care Excellence

Professional organisation submission

Empagliflozin for treating chronic heart failure with reduced ejection fraction [ID3826]

Thank you for agreeing to give us your organisation's views on this technology and its possible use in the NHS.

You can provide a unique perspective on the technology in the context of current clinical practice that is not typically available from the published literature.

To help you give your views, please use this questionnaire. You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.

Information on completing this submission

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 13 pages.

About you	
1. Your name	[redacted]
2. Name of organisation	**The British Society for Heart Failure (BSH)**

3. Job title or position	[redacted]
4. Are you (please tick all that apply):	☒ an employee or representative of a healthcare professional organisation that represents clinicians? ☒ a specialist in the treatment of people with this condition? ☐ a specialist in the clinical evidence base for this condition or technology? ☐ other (please specify):
5a. Brief description of the organisation (including who funds it).	The British Society for Heart Failure (BSH) is a charitable organisation promoting heart failure awareness and education amongst health professionals, patients and policy organisations. Funding is received via the membership and corporate donations.
4b. Has the organisation received any funding from the manufacturer(s) of the technology and/or comparator products in the last 12 months? [Relevant manufacturers are listed in the appraisal matrix.]	£53,000 sponsorship received from Boehringer-Ingelheim comprising; - £8,000 + VAT sponsorship of the BSH annual conference November 2020 - £45,000 grant for Allied Health Professionals eLearning Modules May 2021 (payment still pending) £90,000 sponsorship received from Astra Zeneca comprising; - £60,000 + VAT sponsorship of the BSH annual conference November 2020 - £20,000 + VAT sponsorship for BSH webinar July 2020 - £10,000 + VAT sponsorship as 'Friend' of the BSH 2020-2021

If so, please state the name of manufacturer, amount, and purpose of funding.	
5c. Do you have any direct or indirect links with, or funding from, the tobacco industry?	**None**
The aim of treatment for this condition	
6. What is the main aim of treatment? (For example, to stop progression, to improve mobility, to cure the condition, or prevent progression or disability.)	Heart failure is chronic, progressive condition associated with significant exercise limitation, impaired quality of life, high rates of unplanned hospitalisation and mortality rates comparable to most common forms of cancer. Heart failure therapies therefore aim to delay, prevent or even reverse disease progression thereby resulting in improvements in symptoms, quality of life, hospitalisation and prognosis.

7. What do you consider a clinically significant treatment response? (For example, a reduction in tumour size by x cm, or a reduction in disease activity by a certain amount.)	Clinically significant treatment responses include statistically significant improvements in quality of life endpoints and reducing the risk of hospitalisation and mortality. In the EMPEROR-Reduced trial, the number needed to treat (NNT) to prevent one primary endpoint event (cardiovascular death or heart failure hospitalisation) with Empagliflozin compared to placebo was 19. Whilst there is no agreed optimal NNT for cardiovascular therapies, this number is comparable to the NNT for similar endpoints in other contemporary heart failure clinical trials of approved treatments including; - Sacubitril-Valsartan in the PARADIGM-HF trial, NNT = 21 - Dapagliflozin in the DAPA-HF trial, NNT = 20
8. In your view, is there an unmet need for patients and healthcare professionals in this condition?	Heart failure is a leading cause of hospitalisation and death, with many patients being diagnosed late in their illness and therefore being denied early access to life-extending and life-improving therapies. Even once diagnosed, access to specialist care can be limited. We, the BSH, firmly believe there are significant unmet needs for patients with heart failure and health professionals managing these patients in both primary and secondary care. These unmet needs include high mortality rates, high rates of unplanned hospitalisations and impaired quality of life.
What is the expected place of the technology in current practice?	
9. How is the condition currently treated in the NHS?	As per NICE guidance, patients with suspected heart failure should undergo measurement of blood NT-proBNP and referral to specialist heart failure services for diagnosis and institution of disease modifying therapies. Loop diuretics are given to treat fluid congestion and titrated according to symptoms while standard of care for heart failure with reduced ejection fraction (HFrEF, EF<40%) includes treatment with; 1. a renin-angiotensin-aldosterone-system inhibitor (RAASi) which includes an angiotensin converting enzyme inhibiter (ACEi), angiotensin receptor blocker (ARB) or angiotensin receptor neprilysin

	inhibitor (ARNI) of which, international consensus documents preference ARNI as the RAASi of choice given the superior benefit compared to ACEi/ARBs. 2. a beta-blocker (BB). 3. a mineralocorticoid receptor antagonist (MRA). 4. Dapagliflozin for patients with symptomatic HFrEF despite ARNI/ACEi/ARB, BB and MRA. 5. Ivabradine for patients with symptomatic EF <35% and sinus rhythm heart rate 75 beats per minute or more despite BB, ACEi and MRA. Patients with EF <35% despite optimised medically therapy should be considered for device therapy (cardiac resynchronisation therapy and/or implantable cardioverter defibrillators).

• Are any clinical guidelines used in the treatment of the condition, and if so, which?	There are well-established national and international clinical guidelines in the treatment of heart failure. Two commonly referred to guidelines in England include; 1. NICE clinical guideline 106. Chronic heart failure (2018). 2. European Society of Cardiology guideline for the diagnosis and treatment of acute and chronic heart failure (2016). Please note this guideline is being updated in 2021 and will include recommendation for sodium-glucose transporter-2 inhibitors (SGLT2i). There are also currently four NICE technology appraisals for specialist HF therapies; 1. Ivabradine for treating chronic heart failure [TA267] 2. Implantable cardioverter defibrillators and cardiac resynchronisation therapy for arrhythmias and heart failure [TA314] 3. Sacubitril valsartan for treating symptomatic chronic heart failure with reduced ejection fraction [TA388] 4. Dapagliflozin for treating chronic heart failure with reduced ejection fraction [TA679] In addition, there have been recent consensus statements from other expert specialist bodies reviewing the use and positions of heart failure therapies, including SGLT2i; 1. The American College of Cardiology 2021 Update to the 2017 ACC Expert Consensus Decision Pathway for Optimization of Heart Failure Treatment (https://www.jacc.org/doi/pdf/10.1016/j.jacc.2020.11.022) 2. The UK Cardio-Renal-Metabolic (CaReMe) modified heart failure algorithm (https://www.britishcardiovascularsociety.org/__data/assets/powerpoint_doc/0034/28996/CaReMeUK-HF-March-2021-Final.pptx) designed by a partnership comprising the British Cardiovascular Society, the Renal Association and the Association of British Clinical Diabetologist (https://www.britishcardiovascularsociety.org/resources/bcs-videos-and-webcasts/careme)

• Is the pathway of care well defined? Does it vary or are there differences of opinion between professionals across the NHS? (Please state if your experience is from outside England.)	Diagnostic and treatment pathways for patients with or suspected of having heart failure are well defined in published guidelines (as above) although there are regional/local variations in access to diagnostic tests and interpretation/implementation of some elements of the guidelines.
• What impact would the technology have on the current pathway of care?	Empagliflozin significantly improves symptoms, reduces hospitalisations and reduces cardiovascular mortality in patients with stable chronic heart failure established on standard of care disease-modifying pharmacotherapy (RAASi, BB and MRA) and device therapy, if indicated. Empagliflozin is therefore expected to have significant beneficial impacts on the treatment of patients with HFrEF as add-on therapy to standard of care.
10. Will the technology be used (or is it already used) in the same way as current care in NHS clinical practice?	Empagliflozin is not currently licenced for use in HFrEF but an alternative drug from the same class (Dapagliflozin) is licenced for this use and approved by NICE (NICE TA679, 2021). Dapagliflozin is now being used as add-on therapy to standard of care in patients with symptomatic HFrEF.
• How does healthcare resource use differ between the technology and current care?	
• In what clinical setting should the technology	Empagliflozin is a SGLT2i currently licenced for use in type 2 diabetes mellitus (T2DM) and is well established in primary and secondary care services across the UK. We envisage that Empagliflozin in

be used? (For example, primary or secondary care, specialist clinics.)	HFrEF will be used on the recommendation of a heart failure specialist but could be commenced in primary and secondary care services as it is already well-established in these arena for other purposes.
• What investment is needed to introduce the technology? (For example, for facilities, equipment, or training.)	The initiation of Empagliflozin in HFrEF will require the input from specialist heart failure multi-disciplinary teams and access to blood pressure and blood test monitoring. These investigations form part of standard care for patients with heart failure and since these facilities are already well established across much of the UK, little additional investment is required to introduce Empagliflozin into clinical practice for patients with HFrEF. A small number of additional visits to heart failure specialist teams may also be required, although Empagliflozin requires no dose titration. As such potential visits will represent a small increase to the visits already required. In patients with concomitant T2DM, collaboration with diabetes specialist teams may be necessary and additional training for heart failure specialists in the management of T2DM glucose-lowering agents.
11. Do you expect the technology to provide clinically meaningful benefits compared with current care?	The EMPEROR-Reduced clinical trial clearly demonstrates that in addition to standard of care, compared to placebo, Empagliflozin significantly reduces hospitalisation for heart failure, decline in renal function and improves quality of life. These are all clinically meaningful end-points for patients with heart failure and Empagliflozin is expected to provide significant benefit to these patients.
• Do you expect the technology to increase length of life more than current care?	The EMPEROR-Reduced trial was not sufficiently powered to look at all-cause mortality outcome. Whilst a trend towards benefit in favour of Empagliflozin was demonstrated, this did not reach statistical significance (HR 0.92, 95% CI 0.77-1.10). However, a meta-analysis of the EMPEROR-Reduced and DAPA-HF (a randomised controlled trial of Dapagliflozin in HFrEF) trials did demonstrate a 13% relative risk reduction in all-cause mortality with SGLT2i in HFrEF (HR 0.87, 95% CI 0.77-0.98, p=0.018). Empagliflozin may therefore be inferred to increase length of life in patients with HFrEF on standard of care.
• Do you expect the technology to increase	The EMPEROR-Reduced trial demonstrated significant improvements in quality of life scores as assessed by the Kansas City Cardiomyopathy Questionnaire (KCCQ). A change of 5 points or more is considered

health-related quality of life more than current care?	clinically meaningful and the score change for Empagliflozin was 5.8 vs 4.1 for placebo (HR 1.7, 95% CI 0.5-3). Therefore, Empagliflozin is expected to provide significant quality of life benefits above standard of care.
12. Are there any groups of people for whom the technology would be more or less effective (or appropriate) than the general population?	Empagliflozin has been studied in adult patients with symptomatic HFrEF (EF <40%) and elevated NT-proBNP on standard of care including RAASi, BB, MRA and device therapy if indicated. As with all contemporary heart failure trials, the patient population in EMPEROR-Reduced was younger (mean age 66 years) than average patients with HF in the UK (mean age 77 years). However, this is a consistent feature across all clinical trials of HF therapies, including first-line treatments (ACEi, BB and MRA)

	and specialist treatments such as ivabradine, sacubitril-valsartan, dapagliflozin, cardiac resynchronisation therapy and implantable defibrillators. Furthermore, patients excluded from the EMPEROR-Reduced trial and therefore those in whom benefit is unclear includes; 1. Paediatric patients with HFrEF 2. Patients in NYHA Class I 3. Patients with normal NT-proBNP 4. Hospitalised patients with decompensated heart failure 5. Patients with a heart transplant or recent (within last 90 days) myocardial infarction, major cardiovascular surgery including coronary bypass, stroke or TIA 6. Patients with systolic BP <100 mmHg or >180 mmHg 7. Patients with eGFR <20 ml/min/1.73m2 8. Patients who are pregnant or breast-feeding Whilst patients with Type 1 diabetes mellitus (T1DM) were not excluded from the EMPEROR-Reduced trial, there were none of these patients recruited into the trial and therefore benefit in T1DM is also unclear.
The use of the technology	
13. Will the technology be easier or more difficult to use for patients or healthcare professionals than current	SGLT2i are already well established in current clinical care for use in patients with T2DM. Therefore, transition into patients with HFrEF is expected to be uncomplicated for healthcare professionals. Monitoring for most patients will be in line with usual care for patients with HFrEF although patients with HFrEF and

care? Are there any practical implications for its use (for example, any concomitant treatments needed, additional clinical requirements, factors affecting patient acceptability or ease of use or additional tests or monitoring needed.)	T2DM may require adjustment of other glucose lowering medications. Guidelines for these adjustments are already well established in the diabetic arena.
14. Will any rules (informal or formal) be used to start or stop treatment with the technology? Do these include any additional testing?	Patients with HFrEF are likely to be selected for treatment with Empagliflozin based on current diagnostic pathways that already include NT-proBNP, renal function and echocardiography. Additional testing is not expected for most patients with HFrEF. Patients with concomitant T2DM may require a period of additional glucose monitoring to guide adjustments to other glucose-lowering medications.
15. Do you consider that the use of the technology will result in any substantial health-related benefits that are unlikely to be included in	No

the quality-adjusted life year (QALY) calculation?	
16. Do you consider the technology to be innovative in its potential to make a significant and substantial impact on health-related benefits and how might it improve the way that current need is met?	Empagliflozin joins a number of other SGLT2i's in demonstrating significant outcome benefits in patients with HFrEF. Therefore, whilst the EMPEROR-Reduced data may not be innovative, it maintains potential to provide significant health benefits in patients with HFrEF by adding to the therapy options and improving access to specialist HF therapies.
• Is the technology a 'step-change' in the management of the condition?	The EMPEROR-Reduced trial data adds significant weight to the body of evidence demonstrating significant benefits of SGLT2i's in patients with HFrEF and the SGLT2i class represents a major addition to HF disease-modifying therapy.
• Does the use of the technology address any particular unmet need of the patient population?	Empagliflozin improves morbidity, mortality and quality of life in patients with HFrEF thereby addressing the three core areas of unmet need already described.
17. How do any side effects or adverse effects of the technology affect the	Empagliflozin was well tolerated in the EMPEROR-Reduced trial with overall adverse event rates numerically lower with Empagliflozin (76.2%) than placebo (78.5%) including serious adverse events (41.4% vs 48.1%, respectively. The only excess side-effect noted compared to placebo was uncomplicated genital

management of the condition and the patient's quality of life?	tract infections. These are unlikely to impact on the patients' heart failure status but may impair quality of life if infections are recurrent. Cessation of Empagliflozin resolves any genital tract infection issues.
Sources of evidence	
18. Do the clinical trials on the technology reflect current UK clinical practice?	1/3 of patients in the EMPEROR-Reduced trial were derived from Europe. The standard of care in the trial represents optimal heart failure therapy as practiced within the UK. Although the population in the trial is younger and of a different ethnic makeup to the UK HF population, the results of the trial are broadly applicable to the UK population and UK clinical practice.
• If not, how could the results be extrapolated to the UK setting?	
• What, in your view, are the most important outcomes, and were they measured in the trials?	The EMPEROR-Reduced trial and the meta-analysis of SGLTi's in heart failure, along with the predecessor trials of Empagliflozin in T2DM (EMPA-REG OUTCOME), have all addressed the major outcomes relevant to unmet needs in HF management including; unplanned hospitalisation, mortality and symptoms/quality of life.
• If surrogate outcome measures were used, do they adequately predict long-term clinical outcomes?	The major landmark trials of Empagliflozin have used clinically relevant, hard endpoints rather than surrogate markers to measure clinical outcomes.

• Are there any adverse effects that were not apparent in clinical trials but have come to light subsequently?	None
19. Are you aware of any relevant evidence that might not be found by a systematic review of the trial evidence?	None
20. How do data on real-world experience compare with the trial data?	We are not aware of any currently published data on real-world use of Empagliflozin in HFrEF as it is yet to be licenced for this use.
Equality	
21a. Are there any potential equality issues that should be taken into account when considering this treatment?	The EMPEROR-Reduced trial mainly recruited men (76%) of white background (71%). This is consistent with all major landmark heart failure trials and HFrEF is more common in males in the real world. However, there is no reason to restrict Empagliflozin use in adults based on age or ethnic background.
21b. Consider whether these issues are different from	N/A

issues with current care and why.	

Key messages

22. In up to 5 bullet points, please summarise the key messages of your submission.

- Heart failure is a major global health problem, consuming vast volumes of health resources, particular in relation to a very high burden of unplanned hospitalisations, and causes significant impairment in quality of life for patients. Therapies to combat hospitalisations, quality of life and prognosis are of paramount importance in the long-term management of heart failure.
- Empagliflozin has been comprehensively studied in a randomised, placebo-controlled trial (EMPEROR-Reduced) in patients with symptomatic HFrEF on optimised standard of care including RAASi, BB, MRA and device therapy if indicated, and demonstrates 25% RRR in cardiovascular death or heart failure hospitalisation.
- Empagliflozin significantly improves quality of life measures in patients with HFrEF compared to placebo.
- Empagliflozin is well tolerated, with an acceptable side-effect profile and is easy to initiate and accommodate into existing heart failure services with no need for titration in the treatment of HFrEF.
- The SGLT2i class of medication represent a significant step-change in the management of HFrEF

Thank you for your time.

Please log in to your NICE Docs account to upload your completed submission.

……

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

.....

Empagliflozin for treating chronic heart failure with reduced ejection fraction

Single Technology Appraisal

Source of funding

This report was commissioned by the NIHR Evidence Synthesis Programme as project number 13/48/48T.

Title:	Empagliflozin for treating chronic heart failure with reduced ejection Fraction
Produced by:	BMJ Technology Assessment Group (BMJ TAG)
Authors:	Steve Edwards, Director of Health Technology Assessment, BMJ-TAG, London Victoria Wakefield, Principal Clinical Evidence Analyst, BMJ-TAG, London Mariana Bacelar, Principal Health Economist, BMJ-TAG, London Ben Mayer, Senior Clinical Evidence Analyst, BMJ-TAG, London Charlotta Karner, Clinical Evidence Manager, BMJ-TAG, London
Correspondence to:	Steve Edwards, BMJ TAG, BMJ, BMA House, Tavistock Square, London, WC1H 9JR.
Date completed:	02/09/2021
Source of funding:	This report was commissioned by the NIHR Evidence Synthesis Programme as project number 13/48/48T.
Declared competing interests of the authors	No competing interests were declared which affect the impartiality of this report. BMJ Technology Assessment Group (BMJ-TAG) and the editorial team of The BMJ work independently to one another. The views and opinions expressed in this report are those of the BMJ-TAG.
Acknowledgments:	The ERG would like to thank Dr James Gamble (consultant cardiologist, Oxford University Hospitals NHS Foundation Trust) and Dr Christopher Cassidy (consultant cardiologist, Blackpool Teaching Hospitals NHS Trust) for providing clinical advice throughout the project, and Dr Neil Maredia (consultant cardiologist, South Tees Hospitals NHS Foundation Trust) and Dr Robin Ray (consultant in heart failure, St George's Hospital London) for providing feedback on the clinical sections of the report. The ERG would also like to thank Sarah Roberts (Senior health economist, BMJ-TAG) for providing input on the health economic sections of the report.
Rider on responsibility for report:	The views expressed in this report are those of the authors and not necessarily those of the NIHR Evidence Synthesis Programme. Any errors are the responsibility of the authors.
Report reference:	Edwards SJ, Wakefield V, Bacelar M, Mayer B, Karner C. Empagliflozin for treating chronic heart failure with reduced ejection fraction: A Single Technology Appraisal. BMJ Technology Assessment Group, 2021.

Contribution of authors:

Steve Edwards	Critical appraisal of the company's submission; validated the statistical analyses; provided feedback on all versions of the report. Guarantor of the report
Victoria Wakefield	Critical appraisal of the company's submission; critical appraisal of the clinical evidence; cross checking of company's search strategies; and drafted the summary, background and clinical results sections
Ben Mayer	Critical appraisal of the company's submission; critical appraisal of the clinical evidence; cross checking of company's search strategies; and assisted with drafting the clinical results sections
Charlotta Karner	Critical appraisal of the company's submission; critical appraisal of the clinical evidence; and provided feedback on the clinical sections
Mariana Bacelar	Critical appraisal of the company's submission; critical appraisal of the economic model; cross checking of company's search strategies; critical appraisal of the economic evidence; carried out the economic analyses; and drafted the economic sections

All authors read and commented on draft versions of the ERG report.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACCF	American college of cardiology foundation
ACE	Angiotensin-converting enzyme
ACEi	Angiotensin-converting enzyme inhibitor
AE	Adverse event
AESI	Adverse event of special interest
AF	Atrial fibrillation
AHA	American heart association
AICc	Akaike Information Criterion with a correction for finite sample size
ARB	Angiotensin receptor blockers
ARNi	Angiotensin receptor-neprilysin inhibitor
BB	Beta-blocker
BIC	Bayesian Information Criterion
BNP	B-type natriuretic peptide
CEA	Cost-effectiveness analysis
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence interval
CKD	Chronic kidney disease
CKD-EPI	Chronic kidney disease - Epidemiology collaboration equation
CMR	Cardiac magnetic resonance
COPD	Chronic obstructive pulmonary disease
CPET	Cardiopulmonary exercise test
CQ	Clarification questions
CRM	Cardio-renal-metabolic
CRS	Cardiorenal syndrome
CS	Company submission
CSR	Clinical study report
CT	Computerized tomography
CV	Cardiovascular
DM	Diabetes mellitus
ECG	Electrocardiogram
eGFR	Estimated glomerular filtration rate
EF	Ejection fraction
ESC	European Society of Cardiology
EMA	European medicines agency
EQ-5D	EuroQol- 5 dimension
ERG	Evidence review group

FCE	Finished Consultant Episodes
GEE	Generalised estimating equations
HbA1c	Glycated haemoglobin
HFrEF	Heart failure with reduced ejection fraction
HHF	Hospitalisation for heart failure
HR	Hazard ratio
HRG	Healthcare Resource Group
HRQoL	Health-related quality of life
ICER	Incremental cost-effectiveness ratio
IHD	Ischaemic heart disease
ITC	Indirect treatment comparison
ITT	Intention-to-treat
KCCQ	Kansas City Cardiomyopathy Questionnaire
KCCQ-CSS	Kansas City Cardiomyopathy Questionnaire clinical summary score
KCCQ-TSS	Kansas City Cardiomyopathy Questionnaire total symptom score
KCCQ-OSS	Kansas City Cardiomyopathy Questionnaire overall summary score
KM	Kaplan-Meier
LLA	Lower limb amputation
LVEF	Left ventricular ejection fraction
MedDRA	Medical dictionary for regulatory activities
MI	Myocardial infarction
MMRM	Mixed model for repeated measures
MOA	Mechanism of action
MQT	Market quarter
MRA	Mineralocorticoid receptor antagonists
MHRA	Medicines and Healthcare products Regulatory Agency
MRI	Magnetic resonance imaging
MTD	Maximum tolerated dose
NICE	National Institute for Health and Care Excellence
NT-pro-BNP	N-terminal pro hormone B-type natriuretic peptide
NYHA	New York heart association
OS	Overall survival
PH	Proportional hazard
PT	Preferred term
QALY	Quality adjusted life year
QoL	Quality of life
RCT	Randomised controlled trial

rEF	Reduced ejection fraction
RS	Randomised set
RWE	Real world evidence
SAE	Serious adverse event
SGLT1	Sodium-glucose co-transporter-1
SGLT2	Sodium-glucose co-transporter-2
SGLT2i	Sodium-glucose co-transporter-2 inhibitor
SmPC	Summary of medicinal product characteristics
SoC	Standard of care
T2DM	Type 2 diabetes mellitus
TS	Treated set
TTD	Time to treatment discontinuation
UK	United Kingdom

1 Executive summary

This summary provides a brief overview of the key issues identified by the Evidence Review Group (ERG) as being potentially important for decision making. It also includes the ERG's preferred assumptions and the resulting impact on the incremental cost-effectiveness ratios (ICERs).

Section 1.1 provides an overview of the key issues. Section 1.2 provides an overview of key model outcomes and the modelling assumptions that have the greatest effect on the ICER. Section 1.3 explains the key issues in more detail. Background information on the condition, technology and evidence and information on non-key issues are in the main ERG report.

All issues identified represent the ERG's view, not the opinion of NICE.

1.1 Overview of the ERG's key issues

Table 1. Summary of key issues

ID	Summary of issue	Report sections
1	Uncertainty around the generalisability of the results from EMPEROR-Reduced to the older HFrEF population expected in clinical practice	3.3.10.1
2	Uncertainty around the difference in efficacy of empagliflozin compared with SoC in the Europe subgroup of EMPEROR-Reduced	3.3.10.2
3	Uncertainty around the efficacy of empagliflozin compared with dapagliflozin	3.4
4	The modelling of patients' distribution across the KCCQ-CSS health states	4.1.6.1, 4.1.6.2
5	Use of a Poisson model to estimate HHF	4.1.6.3, 4.1.6.4
6	Overestimation of HHF in the UK population	4.1.6.3, 4.1.6.4
7	Modelling of mortality	4.1.6.8
8	Overestimation of mortality in the UK population	4.1.6.8
9	Impact of HHF in patients' quality of life	4.1.8
10	QoL regressions for the UK population	4.1.8
11	Sex distribution underlying utility estimates	4.1.8
12	Quality of life gains in EMPEROR-R	4.1.8

Abbreviations: HFF, hospitalisation for heart failure; HFrEF, heart failure with reduced ejection fraction; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire clinical summary score; QoL, quality of life; Soc, standard of care; UK, United Kingdom.

The key differences between the company's preferred assumptions and the ERG's preferred assumptions are the fact that the ERG considers that the cost-effectiveness of empagliflozin has not been properly assessed in a population representative of UK clinical practice; and the estimation of

patients' distribution across Kansas City Cardiomyopathy Questionnaire clinical summary score (KCCQ-CSS) states in the model.

1.2 Overview of key model outcomes

National Institute for Health and Care Excellence (NICE) technology appraisals compare how much a new technology improves length (overall survival) and quality of life in a quality-adjusted life year (QALY). An ICER is the ratio of the extra cost for every QALY gained.

Overall, the technology is modelled to affect QALYs by:

- Increasing the proportion of patients who remain in the better KCCQ-CSS states, which in its turn leads to better survival and lower hospitalisation rates.
- Decreasing the probability of patients being hospitalised for heart failure.

Overall, the technology is modelled to affect costs by:

- Its higher unit cost compared to standard of care (SoC) alone.
- Decreasing the probability of patients being hospitalised for heart failure.
- Decreasing the probability of patients receiving dialysis.

The modelling assumptions that have the greatest effect on the ICER are:

- The distribution of patients across the KCCQ-CSS states of the model.
- The rate of hospitalisations for heart failure.
- The cost of dialysis.

1.3 The clinical and cost-effectiveness evidence: summary of the ERG's key issues

The ERG's key issues on the clinical and cost-effectiveness evidence are given in Table 2 to Table 11.

Table 2. Issue 1. Uncertainty around the generalisability of the results from EMPEROR-Reduced to the older HFrEF population expected in clinical practice

Report section	3.3.10.1
Description of issue and why the ERG has identified it as important	The population in EMPEROR-Reduced (hereafter referred to as EMPEROR-R) comprised of patients with more severe HFrEF than would be expected in clinical practice as a result of the inclusion criteria. In addition, the trial population had a mean age of ~67 years which the ERG's clinical experts reported was approximately 10 years younger than the patients they would expect in clinical practice. The ERG notes from the age subgroup analyses in EMPEROR-R that there may be [redacted] In particular, the ERG notes that the benefit

	with empagliflozin compared to placebo in [redacted] The ERG considers it important to highlight that EMPEROR-R wasn't powered to detect differences in treatment effectiveness for the age subgroups and that the results of the subgroups should be interpreted with caution. As per the ERG's request, the company conducted a scenario analysis where the data from the subgroup analysis from EMPEROR-R for patients ≥65 years (mean age 74 years) were used. This scenario used the subgroup analysis undertaken by the company to estimate the HHF and mortality risk equations in the model, together with using a Weibull model as requested by the ERG. As acknowledged by the company, although the results of the subgroup analysis pertain to a population with a similar mean age as that of the patients expected to be treated in the UK, it still reflects a higher risk population than PULSE and the UK population, given the sicker patients included in EMPEROR-R. The ERG remains concerned that the company has not provided a scenario analysis which reflects the whole population considered in this appraisal as the company's subgroups analysis still considerably overestimates HHF in the model when compared to PULSE (and therefore with the UK population).
What alternative approach has the ERG suggested?	The ERG considers baseline characteristics for the age subgroups are required to explore how representative the ≥65 years subgroup is of the HFrEF population expected to be eligible for empagliflozin in UK clinical practice. The ERG recommends that the company undertakes a scenario analysis where HHF KM data from the ≥65 subgroup in EMPEROR-R is used to model time to HHF in the UK population and adjusts the extrapolated curves to reflect a lower number of total HHFs in the model (based on the HHF predictions from PULSE) – see Issue 6. Using the KM HHF data from the EMPEROR-R subgroup would likely still result in an overestimation of HHF when compared to the PULSE population, given the trial inclusion of sicker patients. However, adjusting HHF events in the model to reflect lower hospitalisations in both treatment arms would potentially be easier through the use of extrapolated curves, as these could be adjusted for example, with the use of a HR. Given the availability of KM CV and non-CV mortality data from PULSE, the ERG advises the company to adjust the KM curves from EMPEROR-R to more closely reflect the mortality curves in PULSE – see Issue 8.
What is the expected effect on the cost-effectiveness estimates?	Not predictable, however, it is expected that for the UK subgroup analysis, the overall number of hospitalisations and CV deaths in the model will reduce. With the reduction of overall HHF and CV deaths, the number of events to be avoided with empagliflozin will also decrease, therefore likely increasing the final ICER.
What additional evidence or analyses might help to resolve this key issue?	Full baseline characteristics for each of the age subgroup analyses of EMPEROR-R to establish if there is a baseline difference that might explain the difference in treatment effectiveness across the age subgroups and to see how well aligned the ≥65 years subgroup is with the HFrEF population in UK clinical practice.

Abbreviations: CV, cardiovascular; EMPEROR-R, EMPEROR-Reduced; ERG, evidence review group; HHF, hospitalisation for heart failure; HR, hazard ratio; HFrEF, heart failure with reduced ejection fraction; ICER, incremental cost-effectiveness

ratio; ITT,intention-to-treat; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire clinical summary score; KM, Kaplan-Meier; UK, United Kingdom.

Table 3. Issue 2. Uncertainty around the difference in efficacy of empagliflozin compared with SoC in the Europe subgroup of EMPEROR-Reduced

Report section	3.3.10.2
Description of issue and why the ERG has identified it as important	The ERG is concerned that there are differences in the baseline characteristics of the ITT population of EMPEROR-R compared to patients likely to be eligible for empagliflozin in clinical practice in the UK in terms of age and background use of ACEi/ARBs. Additionally, the ERG is concerned that there appears to be a reduction in efficacy with empagliflozin compared to placebo in the Europe geographical region subgroup analyses from EMPEROR-R. The ERG appreciates that the study was not powered to detect statistically significant differences in subgroups but notes that the Europe subgroup is relatively large (n=1,353). From reviewing the available baseline characteristics of the Europe subgroup, the ERG considers that it may comprise a more severe subgroup of patients compared to the ITT population (and UK population) due to the higher baseline NT-proBNP, and also notes a higher baseline use of implantable cardioverter-defibrillators. The ERG is unclear as to the potential rationale for the observed differences in efficacy in the Europe subgroup and given that this subgroup is of direct relevance to the UK the ERG considers it important to explore further.
What alternative approach has the ERG suggested?	The ERG considers results for all-cause mortality, KCCQ-CSS and renal function for the Europe subgroup should be presented.
What is the expected effect on the cost-effectiveness estimates?	Given the reduced relative treatment effectiveness of empagliflozin observed in this population, it is likely that the ICER will increase.
What additional evidence or analyses might help to resolve this key issue?	Results for all-cause mortality, KCCQ-CSS and renal function for the Europe subgroup.

Abbreviations: ACEi, angiotensin-converting enzyme inhibitor; ARBs, angiotensin receptor blockers; EMPEROR-R, EMPEROR-Reduced; ERG, evidence review group; ICER, incremental cost-effectiveness ratio; ITT,intention-to-treat; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire clinical summary score; UK, United Kingdom.

Table 4. Issue 3. Uncertainty around the efficacy of empagliflozin compared with dapagliflozin

Report section	3.4
Description of issue and why the ERG has identified it as important	There is an absence of head-to-head trial data for the comparison of empagliflozin with dapagliflozin and the company has made an assumption of equal equivalence based on the results of a Bucher ITC and a pooled meta-analysis comprising of only a single trial for each intervention (EMPEROR-R for empagliflozin and DAPA-HF for dapagliflozin). The ERG is concerned that the results of the Bucher ITC show a trend suggesting [redacted] CV deaths and all-cause mortality with empagliflozin compared to dapagliflozin (HR [empagliflozin vs dapagliflozin] [redacted], respectively) and a [redacted] mean change in KCCQ-TSS score from baseline at 8 months with dapagliflozin (MD [redacted]) compared with empagliflozin (MD [redacted]). The results from the ITC are not used in the economic model and instead the company has assumed a class effect for SGLT2is. The ERG considers the results of the Bucher ITC to be [redacted] and that the company's assumption of equal effectiveness for empagliflozin and dapagliflozin lacks robustness.

What alternative approach has the ERG suggested?	The ERG considers that empagliflozin and dapagliflozin should be considered as separate treatments in the economic model with the results of the Bucher ITC used to inform their respective treatment effectiveness.
What is the expected effect on the cost-effectiveness estimates?	Not predictable, however the analysis will not be based on a cost comparison, given the potential difference in effectiveness between the two drugs.
What additional evidence or analyses might help to resolve this key issue?	The use of the clinical effectiveness estimates from the company's Bucher ITC in the economic model to inform the treatment effectiveness of dapagliflozin and generate an ICER for dapagliflozin.

Abbreviations: CI, confidence interval; CV, cardiovascular; EMPEROR-R, EMPEROR-Reduced; ERG, evidence review group; HR, hazard ratio; ICER, incremental cost-effectiveness ratio; ITC, indirect treatment comparison; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire clinical summary score; MD, mean difference; SGLT2is, sodium-glucose co-transporter-2 inhibitors.

Table 5. Issue 4. The modelling of patients' distribution across the KCCQ-CSS health states

Report section	4.1.6.1, 4.1.6.2
Description of issue and why the ERG has identified it as important	The trial reported mean KCCQ-CSS values (and changes) over time, while the model only estimated patients' distribution and movements across the four KCCQ-CSS states defined by the company. Therefore, the validation of the model results against the trial is extremely difficult. It is critical that the changes in KCCQ-CSS predicted by the model are validated against the observed data in EMPEROR-R. The ERG is still unsure which dataset from EMPEROR-R was used to derive KCCQ-CSS changes in the model. Nonetheless, the baseline KCCQ-CSS scores observed in EMPEROR-R would place patients in the model in KCCQ-CSS quartile 2 ([redacted]). Crucially, according to the observed changes in mean KCCQ-CSS in EMPEROR-R, SoC patients would only change KCCQ-CSS quartile in the model at week 52, while empagliflozin patients would only change KCCQ-CSS quartile before week 12 in the model. However, in the model there were improvements in patients' KCCQ-CSS quartiles at all time points in both treatment arms. Importantly, the TPs used in month 9+ of the company's model assume that patients have a very small probability of leaving the KCCQ-CSS state they are in at month 8 in the analysis. This translates into a very strong assumption that the changes seen in EMPEROR-R from baseline to week 52 are sustained for approximately 30 years in the model and that the effect of empagliflozin lasts even after treatment discontinuation, as these patients never catch up to SoC patients. Due to the company's model structure, this also impacts the benefits associated with empagliflozin on HHF and mortality, as these outcomes are dependent on patients' distribution across KCCQ-CSS states.
What alternative approach has the ERG suggested?	The company should: 1. Clarify which dataset from EMPEROR-R is being used to estimate the TPs; 2. Provide the data from EMPEROR-R that allowed the estimation of TPs and proportion of patients in each KCCQ-CSS in the model; 3. Produce the TPs observed in EMPEROR-R for the KCCQ-CSS quartiles defined in the model and explain how these relate to the mean changes reported in the trial; 4. Conduct scenario analyses where the effect of empagliflozin seen at month 8 in the model (sustained by the combination of the proportion of patients in the better KCCQ-CSS states in the

	empagliflozin arm at month 8 and the low probability of disease progression for both SoC and empagliflozin arms in month 9+) wanes over time.
What is the expected effect on the cost-effectiveness estimates?	The direct impact is not predictable; however, this should add clarity to the company's approach. It is expected that the waning scenario analysis will increase the final ICER.
What additional evidence or analyses might help to resolve this key issue?	The company should provide the additional analysis requested in Section 1.6.

Abbreviations: EMPEROR-R, EMPEROR-Reduced; ERG, evidence review group; HHF, hospitalisation for heart failure; ICER, incremental cost-effectiveness ratio; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire clinical summary score; SoC, standard of care; TP, transition probability.

Table 6. Issue 5. Use of a Poisson model to estimate HHF

Report section	4.1.6.3, 4.1.6.4
Description of issue and why the ERG has identified it as important	Despite the model's ability to accurately reproduce the number of HHF observed in the trial over 18 months, the ERG remains uncertain if HHFs are accurately estimated in the long-term model for the trial population. The company's assumption that that the overall rate for HHF remains constant over time across HFrEF patients has not been substantiated by clinical data. Given that time to HHF KM data were available from EMPEROR-R, the ERG considers that the company could have used these data to model time to HHF. Using the KM HHF data from EMPEROR-R would have allowed the company to fit a parametric survival curve to the data and extrapolate into the model's time horizon without having to assume a constant rate of HHF and without having to assume a constant treatment effect with empagliflozin. Furthermore, using KM data for time to HHF would have allowed the company to model time to first and subsequent HHF separately. This could be of importance given the results reported in the EMPEROR-R CSR, indicating that time to subsequent HHF was [redacted] in the empagliflozin than in the placebo arm (at 2 years, [redacted] of patients in the empagliflozin arm had experienced a second HHF, while [redacted] of patients had experienced a second event in the placebo arm).
What alternative approach has the ERG suggested?	The ERG recommends that the company undertakes a scenario analysis where HHF KM data from EMPEROR-R is used to model time to HHF in the trial population.
What is the expected effect on the cost-effectiveness estimates?	Not predictable.
What additional evidence or analyses might help to resolve this key issue?	The company should provide the additional analysis requested in Section 1.6.

Abbreviations: ICER, incremental cost-effectiveness ratio, KCCQ-CSS; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire clinical summary scores; HHF, hospitalisation for heart failure

Table 7. Issue 6. Overestimation of HHF in the UK population

Report section	4.1.6.3, 4.1.6.4
Description of issue and why the ERG has identified it as important	The company conducted a scenario analysis where the subgroup data from EMPEROR-R for patients above 65 years were used in the model to try and reflect the lower rates of HHF seen in PULSE and in clinical practice. Nonetheless, the analysis conducted by the company still grossly overestimates HHF in the model when compared to PULSE (the best available estimate of the UK population). In PULSE, there were 16,033 events observed for the 68,780 HFrEF patients over a mean follow-up of 3 years. The ERG compared these estimates from PULSE to the 3-year HHF outcomes in the model and concluded that the SoC arm of the model overestimates the number of HHF by more than double.
What alternative approach has the ERG suggested?	The ERG recommends that the company undertakes a scenario analysis where HHF KM data from the >65 subgroup in EMPEROR-R is used to model time to HHF in the UK population and adjusts the extrapolated curves to reflect a lower number of total HHFs in the model (based on the HHF predictions from PULSE). Using the KM HHF data from the EMPEROR-R subgroup would likely still result in an overestimation of HHF when compared to the PULSE population, given the trial inclusion of sicker patients. However, adjusting HHF events in the model to reflect lower hospitalisations in both treatment arms would potentially be easier through the use of extrapolated curves, as these could be adjusted for example, with the use of a HR.
What is the expected effect on the cost-effectiveness estimates?	Not predictable, however, it is expected that the overall number of hospitalisations in the model will reduce for this population. With the reduction of overall HHF, the number of HHF to be avoided with empagliflozin will also decrease, therefore likely increasing the final ICER.
What additional evidence or analyses might help to resolve this key issue?	The company should provide the additional analysis requested in Section 1.6.

Abbreviations: EMPEROR-R, EMPEROR-Reduced; ERG, evidence review group; HHF, hospitalisation for heart failure; HR, hazard ratio; HFrEF, heart failure with reduced ejection fraction; ICER, incremental cost-effectiveness ratio; KM, Kaplan-Meier; SoC, standard of care; UK, United Kingdom.

Table 8. Issue 7. Modelling of mortality

Report section	4.1.6.8
Description of issue and why the ERG has identified it as important	The ERG considers that the assumption of a constant treatment effect (and therefore PHs) of empagliflozin over SoC throughout the model is unsubstantiated, both for all-cause and CV-related death. The empagliflozin and placebo KM OS curves hardly separate during the follow-up period of the trial and the HRs in EMPEROR-R for all-cause and CV-related death were not statistically significant (and also signalled a small effect size). Given the company's own assessment that there were not enough CV deaths in EMPEROR-R to establish a robust effect of empagliflozin, and that CV-related deaths represented 75.5% of all deaths in EMPEROR-R; the ERG considers that there is not enough evidence to support the inclusion of a treatment effect in CV and non-CV mortality in the economic model.

What alternative approach has the ERG suggested?	The ERG conducted a scenario analysis where CV and non-CV mortality were assumed to be the same in the empagliflozin and the SoC arms. It is important to note that when no treatment effect is assumed for empagliflozin on mortality in the economic model, there is still a benefit associated with empagliflozin on both CV and non-CV mortality. This is because the probability of patients dying is different in every KCCQ-CSS state of the model. Given that patients in the empagliflozin arm of the model have a higher probability of remaining in the better KQCC-CSS states over time compared with SoC patients, the former also experience a lower probability of death. The ERG recommends that the company considers adding a scenario analysis in the model where it is assumed that empagliflozin has no survival benefit over SoC (including through the residency in KCCQ-CSS states).
What is the expected effect on the cost-effectiveness estimates?	For the trial population, assuming that empagliflozin had no effect on CV and non-CV mortality increased the ICER from £4,717 to £5,712. For the UK population, assuming that empagliflozin had no effect on CV and non-CV mortality increased the ICER from £6,342 to £7,270.
What additional evidence or analyses might help to resolve this key issue?	During clarification, the ERG requested that the company provided independently fitted curves to the trial arms in order to extrapolate CV-related deaths in the model. The ERG was concerned that the fitted Weibull model was a poor fit to the underlying CV mortality KM data, particularly to the empagliflozin arm and therefore, recommended that the company provided a more flexible modelling approach. The company did not undertake such analysis because it considered that independently fitted curves led to implausible long-term patterns with projections for placebo leading to longer mean time to CV-death with all the tested distributions. The company added that CV deaths were relatively rare with around 10% of patients dying due to a CV event. If the company can substantiate modelling a treatment effect for empagliflozin on patients' mortality, the ERG recommends that the company adopts a modelling approach where CV mortality and all-cause mortality KM data are fitted independently in the model. The company should also provide the additional analysis requested in Section 1.6.

Abbreviations: CV, cardiovascular; EMPEROR-R, EMPEROR-Reduced; ERG, evidence review group; HR, hazard ratio; ICER, incremental cost-effectiveness ratio; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire clinical summary scores; KM, Kaplan-Meier; OS, overall survival; PH, proportional hazard; SoC, standard of care; UK, United Kingdom.

Table 9. Issue 8. Overestimation of mortality in the UK population

Report section	4.1.6.8
Description of issue and why the ERG has identified it as important	When compared to the PULSE results, the SoC arm of the model overestimates the number of CV-related deaths (and underestimates the number of non-CV deaths) when the subgroup from EMPEROR-R is modelled. This reflects a population more likely to die of CV causes than PULSE patients. The company conducted a scenario analysis where the data for the subgroup analysis from EMPEROR-R for patients above 65 years in the trial were used. The company has not provided the KM data for all-cause or CV mortality in the above 65 years group, therefore the ERG cannot validate the

	newly fitted Weibull regressions against the appropriate KM data from the EMPEROR-R subgroup. In PULSE, there were 7,905 CV deaths and 9,599 non-CV deaths over a mean follow-up of 3 years. In the model, there were [redacted] CV deaths and [redacted] non-CV deaths when the subgroup data from EMPEROR-R is used (for the first 3 years in the model). The ERG remains concerned that the results of the company's analysis on the cost-effectiveness of empagliflozin in the UK population with HFrEF are not reliable.
What alternative approach has the ERG suggested?	The lack of flexibility in the company's model structure and the lack of KCCQ-CSS data from PULSE mean that the company cannot use mortality data from PULSE directly in the SoC arm of the model. However, given the availability of KM CV and non-CV mortality data from PULSE, the ERG considers that this could be done by adjusting the KM curves from EMPEROR-R to more closely reflect the mortality curves in PULSE.
What is the expected effect on the cost-effectiveness estimates?	Not predictable, however, it is expected that for the UK subgroup analysis, the number of CV deaths in the model will reduce. With the reduction of CV deaths, the number of events to be avoided with empagliflozin will also decrease, therefore likely increasing the final ICER.
What additional evidence or analyses might help to resolve this key issue?	The company should supply the KM data for all-cause and CV mortality in the above 65 years group. The company should provide the additional analysis requested in Section 1.6.

Abbreviations: CV, cardiovascular; EMPEROR-R, EMPEROR-Reduced; ERG, evidence review group; HFrEF, heart failure with reduced ejection fraction; ICER, incremental cost-effectiveness ratio; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire clinical summary scores; KM, Kaplan-Meier; SoC, standard of care; UK, United Kingdom.

Table 10. Issue 9. Impact of HHF in patients' quality of life

Report section	4.1.8
Description of issue and why the ERG has identified it as important	The ERG considers that the impact of HHF on patients' quality of life is overestimated in the model. The company arrived at the -0.246 disutility per HHF event by assuming that all HHFs in the model last for 1 year. The ERG considers it unlikely that all HHFs events will have that duration. The ERG notes that the second biggest driver of the QALY gains in the model comes from the reduction in HHF for empagliflozin patients when compared to SoC patients.
What alternative approach has the ERG suggested?	The ERG advises that the company reports the proportion of patients in EMPEROR-R who were hospitalised for 1; 2; and 8 months in the trial and generates a weighted disutility value to be applied in the model. Ideally, for the UK population, the same analysis would be conducted using PULSE data, as the mean duration of HHF is likely to be lower in PULSE than in EMPEROR-R.
What is the expected effect on the cost-effectiveness estimates?	It is expected that the disutility value associated with HHF will decrease, therefore reducing the QALY gain associated with empagliflozin.

What additional evidence or analyses might help to resolve this key issue?	The ERG notes that it did not have access to mean (or median) duration of hospitalisations in EMPEROR-R, and so it cannot ascertain the extent to the overestimation of this disutility. The company should provide as much detail as possible on duration of HHF in EMPEROR-R.

Abbreviations: EMPEROR-R, EMPEROR-Reduced; ERG, evidence review group; HHF, hospitalisation for heart failure; QALY, quality adjusted life year; SoC, standard of care; UK, United Kingdom.

Table 11. Issue 10. QoL regressions for the UK population

Report section	4.1.8
Description of issue and why the ERG has identified it as important	The baseline characteristics from the older EMPEROR-R subgroup were used in the QoL regression analysis, however the regression was not re-estimated in this subgroup and thus the coefficients for the predictors remained the same as those for the ITT population.
What alternative approach has the ERG suggested?	The ERG recommends that the company re-estimates the regression model using the subgroup data.
What is the expected effect on the cost-effectiveness estimates?	Not predictable.
What additional evidence or analyses might help to resolve this key issue?	As above.

Abbreviations: EMPEROR-R, EMPEROR-Reduced; ERG, evidence review group; ITT, intention-to-treat; QoL, quality of life.

1.4 Other key issues: summary of the ERG's view

Table 12. Issue 11. Sex distribution underlying utility estimates

Report section	4.1.8
Description of issue and why the ERG has identified it as important	Trial population: The ERG is concerned that the 0.7740 utility value associated with the KCCQ-CSS quartile 4 state (taken from Sullivan *et al.*)[1] does not accurately reflect the baseline gender distribution in EMPEROR-R. In Sullivan *et al.*[1] the population was composed of 52% females and 48% males. In EMPEROR-R, only 24% of the population were females. UK population: The ERG is concerned that the 0.723 utility value associated with the KCCQ-CSS quartile 4 state (taken from Sullivan *et al.*)[1] does not accurately reflect the baseline gender distribution in the UK population. In PULSE, only 35% of the population were females.
What alternative approach has the ERG suggested?	Trial population: The ERG recommends that the company adjusts the 0.7740 value in the trial population analysis to reflect the gender distribution in EMPEROR-R. UK population: The ERG recommends that the company adjusts the 0.723 value in the trial population analysis to reflect the gender distribution in PULSE.

What is the expected effect on the cost-effectiveness estimates?	The utility values reported by Sullivan *et al*.[1] suggest that males experienced a higher utility value than females. Given that in the study the percentage of females was higher than in EMPEROR-R and PULSE, the ERG expects the adjustments to result in a lower utility value associated with the KCCQ-CSS quartile 4 state for both populations.
What additional evidence or analyses might help to resolve this key issue?	As above.

Abbreviations: EMPEROR-R, EMPEROR-Reduced; ERG, evidence review group; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire clinical summary score; UK, United Kingdom.

Table 13. Issue 12. Quality of life gains in EMPEROR-R

Report section	4.1.8
Description of issue and why the ERG has identified it as important	The ERG is concerned that the EQ-5D data from EMPEROR-R [redacted] in empagliflozin patients' QoL when compared to placebo patients and that the economic model generates a QALY gain of [redacted]. The two main drivers of QALY gain in the model are related to: 1) how much longer empagliflozin patients stay in the better KCCQ-CSS states; and 2) the reduction in HHF experienced by empagliflozin patients.
What alternative approach has the ERG suggested?	For the trial population: The scenario analysis suggested for the KCCQ-CSS modelling (see Issue 4) in combination with the adjustments to the QALY calculations suggested in Issue 9 and Issue 11. For the UK population: A combination of the following scenario analyses: The scenario analysis suggested for the KCCQ-CSS modelling (see Issue 4); the adjustments to the QALY calculations suggested in Issue 9, Issue 10 and Issue 11; the scenario analysis suggested in Issue 6 to reduce the number of HHF in the model; the scenario analysis suggested in Issue 8 to reduce the number of CV deaths in the model
What is the expected effect on the cost-effectiveness estimates?	It is expected that the QALY gain associated with empagliflozin will decrease.
What additional evidence or analyses might help to resolve this key issue?	As above.

Abbreviations: CV, cardiovascular; EMPEROR-R, EMPEROR-Reduced; EQ-5D, EuroQol- 5 dimension; ERG, evidence review group; HHF, hospitalisation for heart failure; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire clinical summary score; QALY, quality adjusted life year; QoL, quality of life; UK, United Kingdom.

1.5 Summary of ERG's preferred assumptions and resulting ICER

The ERG conducted two sets of analysis, one using the entire trial population from EMPEROR-Reduced (EMPEROR-R) and the other using the above 65 years subgroup from the trial. The former analysis aims to estimate the cost-effectiveness of empagliflozin vs SoC in the trial population, while the latter analysis intends to ascertain the cost-effectiveness of empagliflozin vs SoC in the UK population with heart failure with reduced ejection fraction (HFrEF). Nonetheless, the ERG remains

concerned that using the older subgroup of patients from EMPEROR-R still reflects a sicker population than that seen in UK clinical practice, with higher mortality and hospitalisation for heart failure (HHF). The ERG reinforces the need for the company to supply the full baseline characteristics for the ≥65 population in EMPEROR-R to establish the main differences with the HFrEF population in clinical practice in the UK.

Results of the exploratory analyses conducted using the trial population are reported in Table 14, while Table 15 reports the results in the UK population analysis. The following analyses were conducted in both populations:

1. Assuming that 84% of patients receive lifelong treatment with empagliflozin;
2. Using a Weibull model to estimate time to treatment discontinuation (TTD) for empagliflozin;
3. Using the relative utility adjustment and the age-related decrements from Ara;
4. Replacing the proportion of UK patients who receive angiotensin-converting enzyme inhibitor (ACEi) and angiotensin receptor-neprilysin inhibitor (ARNi) in the model to reflect the ERG's clinical experts' opinion;
5. Using a unit cost for cardiovascular (CV) death in the model of £1,582;
6. Applying the ERG-calculated annual cost of dialysis (assuming 3 weekly sessions) of £23,088;

The analyses listed below are specific to each population:

Trial population

a. Assuming that non-CV and CV mortality is the same for empagliflozin and SoC and using a Gompertz curve to estimate mortality.

UK population

b. Assuming that non-CV and CV mortality is the same for empagliflozin and SoC and using a Weibull curve to estimate mortality.
c. Using the 0.723 utility value from Sullivan for the KCCQ-CSS quartile state (and adjusting other KCCQ-CSS state values accordingly).

Results in Table 14 and Table 15 show that the ICERs in both populations do not change by more than approximately £1,000 per QALY gained. This is a direct consequence of the lack of flexibility in

the economic model and it demonstrates how the key clinical outcomes (such as mortality) are intrinsically linked to the distribution of patients across the KCCQ-CSS states.

The ERG could not produce its preferred ICERs for each population given the remaining uncertainty in the long-term effect of empagliflozin on patients' change in KCCQ-CSS (in both the trial and in the UK population analyses sets) and the lack of representativeness of the subgroup data from EMPEROR-R when trying to replicate the UK population. The overestimation of CV mortality and the overestimation of HHF in the model compared to the UK population when the >65 years subgroup is used in the model indicate the lack of external validity of the model results in this population.

The ERG conducted one additional scenario analysis to try and explore these areas of remaining uncertainty. The scenario assumed that empagliflozin had no effect on patients' transitioning through KCCQ-CSS states. This means that SoC and empagliflozin patients were distributed equally across the KCCQ-CSS states in the economic model. When this assumption was used in the economic model (in combination with the ERG's preferred assumptions), the final ICER for the trial population was £15,716, while for the UK population was £31,924 per QALY gained. In both cases, the main driver of QALY gain in the model was the benefit of empagliflozin on HHF.

Given that for the UK population analysis, the SoC arm of the model overestimates the number of HHF by more than double when compared to PULSE, the ERG notes that the ICER of £31,924 is likely to increase substantially if the number of HHF was reduced in the model. The ERG could not robustly adjust the HHF rate in the model as this would entail artificially manipulating the coefficients for the HHF regression to produce a lower number of HHFs in the model. The ERG, therefore, recommends that the company conducts this analysis.

The scenario analysis conducted by the ERG indicates that the ICER for empagliflozin compared to SoC is likely to remain under the £30,000 threshold in the trial population, even when it is assumed that empagliflozin has no effect on patients' movements through the KCCQ-CSS quartiles defined by the company. Nonetheless, the ERG remains concerned that the cost-effectiveness of empagliflozin compared to SoC in the UK population remains highly uncertain.

Table 14. Results of ERG's exploratory analysis in the trial population

	Results per patient	Empagliflozin	SoC	Incremental value
0	Company's base case post clarification			
	Total costs	£17,837	£16,911	£926
	QALYs	3.76	3.56	0.20
	ICER (£/QALY)	-	-	£4,717

1	Assuming that 84% of patients receive lifelong treatment with empagliflozin			
	Total costs	£18,433	£16,911	£1,522
	QALYs	3.86	3.56	0.30
	ICER (£/QALY)	-	-	£5,008
2	Using a Weibull model to estimate TTD for empagliflozin			
	Total costs	£17,893	£16,911	£981
	QALYs	3.77	3.56	0.21
	ICER (£/QALY)	-	-	£4,763
a	Assuming that non-CV and CV mortality is the same for empagliflozin and SoC using a Gompertz curve			
	Total costs	£12,798	£12,194	£604
	QALYs	2.63	2.53	0.11
	ICER (£/QALY)	-	-	£5,712
3	Using the relative utility adjustment and the age-related decrements from Ara			
	Total costs	£17,837	£16,911	£926
	QALYs	3.68	3.49	0.19
	ICER (£/QALY)	--	-	£4,915
4	Replacing the proportion of UK patients who receive ACEi and ARNi in the model to reflect the ERG's clinical experts' opinion			
	Total costs	£17,115	£16,212	£903
	QALYs	3.76	3.56	0.20
	ICER (£/QALY)	-	-	£4,602
5	Using a unit cost for CV death in the model of £1,582			
	Total costs	£16,405	£15,452	£952
	QALYs	3.76	3.56	0.20
	ICER (£/QALY)	-	-	£4,853
6	Applying the ERG-calculated annual cost of dialysis (assuming 3 weekly sessions) of £23,088			
	Total costs	£17,576	£16,565	£1,011
	QALYs	3.76	3.56	0.20
	ICER (£/QALY)	-	-	£5,152

Abbreviations: ACEi, angiotensin-converting enzyme inhibitor; ARNi, angiotensin receptor-neprilysin inhibitor; CV, cardiovascular; ERG, evidence review group; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year; SoC, standard of care; TTD, time to treatment discontinuation.

Table 15. Results of ERG's exploratory analysis in the UK population

	Results per patient	Empagliflozin	SoC	Incremental value
0	Company's base case post clarification using >65 subgroup data from EMPEROR-R + using Weibull curve to estimate mortality			
	Total costs	£16,436	£15,198	£1,238
	QALYs	3.55	3.36	0.20
	ICER (£/QALY)	-	-	£6,342
1	Assuming that 84% of patients receive lifelong treatment with empagliflozin			
	Total costs	£17,134	£15,198	£1,937

	QALYs	3.64	3.36	0.29
	ICER (£/QALY)	-	-	£6,795
2	Using a Weibull model to estimate TTD for empagliflozin			
	Total costs	£16,502	£15,198	£1,305
	QALYs	3.56	3.36	0.20
	ICER (£/QALY)	-	-	£6,410
b	Assuming that non-CV and CV mortality is the same for empagliflozin and SoC using a Weibull curve and the above 65 subgroup population in the model			
	Total costs	£16,168	£15,198	£971
	QALYs	3.49	3.36	0.13
	ICER (£/QALY)			£7,270
3	Using the relative utility adjustment and the age-related decrements from Ara			
	Total costs	£16,436	£15,198	£1,238
	QALYs	3.47	3.28	0.19
	ICER (£/QALY)	-	-	£6,641
c	Using the 0.723 utility value from Sullivan for the KCCQ-CSS quartile 4 state (and adjusting other KCCQ-CSS state values accordingly)			
	Total costs	£16,436	£15,198	£1,238
	QALYs	3.34	3.16	0.18
	ICER (£/QALY)	-	-	£6,758
4	Replacing the proportion of UK patients who receive ACEi and ARNi in the model to reflect the ERG's clinical experts' opinion			
	Total costs	£15,816	£14,602	£1,214
	QALYs	3.55	3.36	0.20
	ICER (£/QALY)	-	-	£6,219
5	Using a unit cost for CV death in the model of £1,582			
	Total costs	£15,092	£13,841	£1,251
	QALYs	3.55	3.36	0.20
	ICER (£/QALY)	-	-	£6,407
6	Applying the ERG-calculated annual cost of dialysis (assuming 3 weekly sessions) of £23,088			
	Total costs	£16,217	£14,907	£1,311
	QALYs	3.55	3.36	0.20
	ICER (£/QALY)	-	-	£6,713

Abbreviations: ACEi, angiotensin-converting enzyme inhibitor; ARNi, angiotensin receptor-neprilysin inhibitor; CV, cardiovascular; EMPEROR-R, EMPEROR-Reduced; ERG, evidence review group; ICER, incremental cost-effectiveness ratio; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire clinical summary score; QALY, quality adjusted life year; SoC, standard of care; TTD, time to treatment discontinuation.

1.6 Additional data request for model validation

Given the difficulties around validating the model KCCQ-CSS results against the trial, discussed in Issue 4, the ERG recommends that the company provides additional data from EMPEROR-R in the

format outlined in Table 16. The ERG notes that the company should complete Table 16 twice; once for the ITT population in EMPEROR-R, and another for the ≥65 years subgroup from the trial.

Table 16. KCCQ-CSS data from EMPEROR-R for model validation

Data	Empagliflozin + SoC			Placebo + SoC		
	Months 1-3	Months 4-8	Months 9+	Months 1-3	Months 4-8	Months 9+
KCCQ 1						
n/N*						
Number of HHF						
Number of CV deaths						
Number of all-cause deaths						
KCCQ 2						
n/N*						
Number of HHF						
Number of CV deaths						
Number of all-cause deaths						
KCCQ 3						
n/N*						
Number of HHF						
Number of CV deaths						
Number of all-cause deaths						
KCCQ 4						
n/N*						
Number of HHF						
Number of CV deaths						
Number of all-cause deaths						

Abbreviations: CV, cardiovascular; EMPEROR-R, EMPEROR-Reduced; HHF, hospitalisation for heart failure; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire clinical summary score.

*n/N – please provide the number of patients in the respective KCCQ-CSS state divided by the total number of patients alive at the same time point

2 Introduction and background

2.1 Introduction

Herein is a critique of the evidence submitted to the Single Technology Appraisal (STA) in support of the clinical and cost effectiveness of empagliflozin (Jardiance®, Boehringer Ingelheim) for treating chronic heart failure with reduced ejection fraction.[2]

2.2 Background

Within Section B.1 of the company submission (CS), the company provides an overview of:

- empagliflozin, including its mode of action, dose and method of administration (CS, Section B.1.2);
- heart failure (HF), including aetiology, classification of HF, prevalence, comorbidities and risk factors for HF, burden of disease and current disease management (CS, Section B.1.3).

Based on advice from its clinical experts, the Evidence Review Group (ERG) considers the CS to present an accurate overview of the aetiology and diagnosis of HF, and the management of the condition. To aid understanding of some points raised in the ERG's critique of the submitted evidence in the context of the decision problem, the ERG provides an overview of key aspects of the management of heart failure with reduced ejection fraction (HFrEF).

HFrEF is a subtype of HF that is categorised as a left ventricular ejection fraction (LVEF) of less than 40% and results in the left side of the heart not pumping blood around the body as well as would be expected. The focus of this STA is on the patient population with symptomatic HFrEF. The New York Heart Association (NYHA) classification[3] is a commonly used tool for functional classification of patients with HF (Table 17), although it should be noted that the classification is performed by a clinician and is a subjective measure of a patient's functional capacity. The ERG's clinical experts reported that patients in Class I would be asymptomatic and those in Class II or above would be symptomatic.

Table 17: NYHA functional classification based on severity of symptoms and physical activity [3] (Reproduced from CS, Table 5)

Classification	Description
Class I	No limitation of physical activity. Ordinary physical activity does not cause symptoms of heart failure
Class II	Slight limitation of physical activity. Comfortable at rest, but ordinary physical activity results in symptoms of heart failure

Class III	Marked limitation of physical activity. Comfortable at rest, but less than ordinary activity causes symptoms of heart failure
Class IV	Unable to carry on any physical activity without symptoms of heart failure, or symptoms of heart failure at rest
Abbreviation: CS, company submission; NYHA, New York Heart Association	

The National Institute for Health and Care Excellence (NICE) guideline on diagnosis and management of chronic heart failure in adults (NG106) includes detailed recommendations for the management of HFrEF.[4] In summary, it recommends a sequential approach to treatment and treatments are broken down into first line and specialist treatments.

The first line treatment recommendations in NG106 comprise an angiotensin-converting enzyme inhibitor (ACEi) (or angiotensin receptor-blocker [ARB] if intolerant to ACEi) and a beta-blocker (BB). Diuretics may also be used to provide symptomatic relief. If symptoms continue, mineralocorticoid receptor antagonists (MRA) should be added to the ACEi or ARB if there is no evidence of hyperkalaemia.

Specialist treatments should only be initiated by a HF specialist if symptoms persist after dose optimisation of standard of care (SoC) therapy with ACEi/ARB, BB and/or MRA combination. The specialist treatments comprise:

- Sacubitril valsartan as an alternative to ACEi or ARB in patients with continuing New York Heart Association (NYHA) class II-IV symptoms and LVEF ≤ 35%;
- Addition of ivabradine to the SoC for patients in sinus rhythm with a heart rate ≥ 75 beats per minute and LVEF ≤ 35%;
- Addition of hydralazine and nitrate especially in patients of African-Caribbean descent with moderate to severe HF, or in patients who can tolerate neither an ACEi nor an ARB; and
- Digoxin in patients with worsening or severe HFrEF with sinus rhythm.

The ERG notes that the company reports there has been a low uptake of ivabradine, hydralazine/nitrate and digoxin and that these treatments are not relevant for all HFrEF patients in which empagliflozin in indicated.[5] Additionally, the ERG notes that this is consistent with the NICE single technology appraisal for dapagliflozin (TA679), which comprises the same population as this STA, and where ivabradine, hydralazine and nitrate, and digoxin were not deemed to be relevant comparators for dapagliflozin.[6] The only one of these specialist drugs from NG106 specified in the

NICE final scope for this appraisal as a comparator of interest is sacubitril valsartan, and the ERG's clinical experts and the company agree with this inclusion.

Dapagliflozin was recently approved (February 2021)[6] for use in adults with symptomatic chronic HFrEF as an add-on to optimised standard care with:

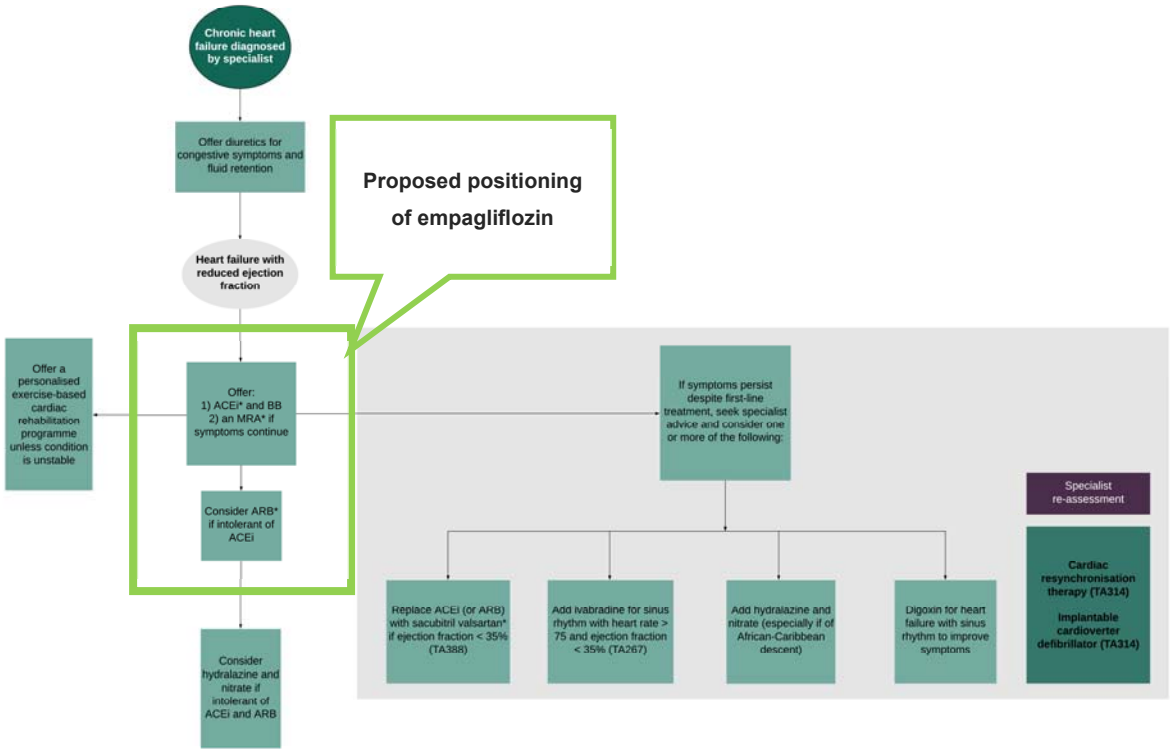
- ACEi or ARBs, with BBs, and, if tolerated, MRAs, or
- sacubitril valsartan, with BBs, and, if tolerated, MRAs.

The ERG notes that the dapagliflozin recommendation requires the involvement of a HF specialist to initiate treatment and that it was not available at the time of publication of NG106. The company argue that dapagliflozin should not be included as a comparator for empagliflozin as it's uptake also remains low, although the ERG and the ERG's clinical experts disagree and consider dapagliflozin use to be increasing and that it is a relevant comparator.

2.2.1 Positioning of empagliflozin in the UK treatment pathway

The company's proposed positioning of empagliflozin in the NICE pathway is, "*as an add-on to ACEi or ARBs plus BB, and/or MRA therapy for HFrEF patients with or without comorbidities, who continue to be symptomatic while receiving stable, but not necessarily optimised doses of SoC*" (Figure 1).

Figure 1. Company proposed positioning of empagliflozin in the HFrEF treatment pathway (Reproduced from CS, Figure 4)



*Measure serum sodium and potassium, and assess renal function before and after starting and after each dose increment. If eGFR is 30 to 45 ml/min/1.73m^2, consider lower doses or slower titration of ACEi or ARBs, MRAs, sacubitril valsartan and digoxin.

Abbreviations: ACEi angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor-blocker; BB, beta-blocker; MRA, mineralocorticoid receptor antagonist

Source: Adapted from NICE guideline NG106, 2018 [4]

The preferred positioning for empagliflozin (in light green box) is in primary care in recently diagnosed symptomatic patients who receive SoC (e.g., ACE, BB, ARB), but not necessarily with optimised dosing

The company considers the relevant comparators for empagliflozin are:

- Individualised standard care defined as:
    - ACEi in combination with a BB, and/or MRA;
    - ARB in combination with a BB, and/or MRA;
    - Sacubitril valsartan in combination with BB, and/or MRA.

The ERG's clinical experts reported that they would anticipate commencement of empagliflozin by, or on the recommendation of a HF specialist similar to dapagliflozin although it could be commenced in either primary or secondary care. The ERG also notes that the company are proposing standard care drugs are not required to be optimised prior to commencement of empagliflozin, although this is not in keeping with the use of empagliflozin in the key study informing the clinical effectiveness data in the CS (EMPEROR-Reduced [hereafter referred to as EMPEROR-R]).

Lastly, as discussed above (Section 2.2), the ERG and its clinical experts disagree with the company's proposal that dapagliflozin is not a relevant comparator. However, the ERG notes that the company has provided results from an ITC of empagliflozin vs dapaglflozin in the CS.

2.3 Critique of the company's definition of the decision problem

The company provided a summary of the final scope issued by the NICE, together with their rationale for any deviation from the final scope (Table 18).[2] The company highlights that the submission differs from the final scope primarily in terms of the comparators of interest to the decision problem. The key differences between the decision problem addressed in the CS and the scope are discussed in greater detail in the sections that follow.

Table 18. Summary of decision problem

	Final scope issued by NICE	Decision problem addressed in the submission	Rationale if different from the scope	ERG comment
Population	Adults for the treatment of symptomatic chronic heart failure with reduced ejection fraction	Same	Not applicable	Evidence derived from EMPEROR-R[7] is aligned with the population specified in the final scope although the ERG's clinical experts report that the baseline characteristics of the trial population are reflective of younger patients with more advanced heart failure compared to the UK patient population who would be eligible for treatment with empagliflozin. In addition, the ERG's experts reported the trial population comprised of more males than expected in clinical practice (more detailed description of population available in Section 2.3.1 and Section 3.2).
Intervention	Empagliflozin in combination with standard care (including diuretics, treatment with an ACE inhibitor, ARBs, mineralocorticoid receptor antagonist, beta blockers, cardiac devices and sacubitril valsartan)	Same	Not applicable	The intervention in EMPEROR-R matches the intervention specified in the final scope, that is empagliflozin in addition to SoC. SoC in EMPEROR-R comprised a guideline-directed medical therapy with ACEi/ARB or ARNi, beta-blockers and mineralocorticoid receptor antagonists. The ERG's clinical experts reported that the use of ACEi would be expected to be slightly higher in clinical practice compared to in EMPEROR-R (approx. 80% instead of

				approx. 70%) and ARNi possibly slightly lower than 20%. More detail on the intervention is provided in Sections 2.3.2 and 3.2.
Comparator(s)	• Individually optimised standard care without empagliflozin. Standard care is defined as: o ACE inhibitors in combination with betablockers, and/or mineralocorticoid receptor antagonists o ARBs in combination with beta-blockers, and/or mineralocorticoid receptor antagonists o Sacubitril valsartan in combination with beta-blockers, and/or mineralocorticoid receptor antagonists • Dapagliflozin as an add on to standard care	Same, however dapagliflozin does not reflect current SoC and is not a relevant comparator. The evidence for empagliflozin vs standard treatment with ACEi, ARBs, sacubitril/valsartan is the most relevant for the committee to consider. This is because a majority of eligible patients in the UK receive at least one of these products. Comparative analyses of empagliflozin vs dapagliflozin are provided in [CS Section B.2.8] upon the request of the ERG and NICE Technical team; however, these are secondary. This is consistent with the perspective of UK clinicians we have consulted.	The estimated prescribing of dapagliflozin is [redacted] in MQT May 2021 for patients with HF only. In the HF only population, it is prescribed [redacted] times less often than sacubitril valsartan [CS, Table 2], a product considered as SoC but prescribed less frequently than ACEi and ARBs[8]. The NICE dapagliflozin resource impact template estimated that 75% of HFrEF patients optimised on standard care with either ACE/ARBs, or sacubitril/valsartan and with an eGFR >30mL/min per 1.73m2 will receive dapagliflozin by 2025. In 2021, 2022, 2023, 2024, the uptake is estimated to be 20%, 35%, 50%, and 60%, respectively.[6] The resource impact template only considers patients with HFrEF only and not those with comorbid T2DM. There is limited empirical evidence to support these estimates. Given the market share in MQT May 2021 was [redacted] it's unlikely that 20% of eligible HFrEF only patients would receive dapagliflozin by the end of 2021.	The comparator in EMPEROR-R was placebo as add on to SoC which the ERG considers appropriate for the comparison of empagliflozin with SoC. The SoC treatments are combined in the clinical analyses which the ERG's clinical experts reported to be reasonable and the ERG notes is in keeping with the approach taken in TA679[6] (dapagliflozin). For the comparison with dapagliflozin the company conducted a Bucher ITC due to the absence of head-to-head trial data. The ERG's clinical experts reported that dapagliflozin uptake is increasing and that it should be considered a key comparator for empagliflozin despite the company's assertions that it is not relevant as it is not routinely used at present. The ERG is concerned that the company is making a strong assumption of equivalence for the clinical-effectiveness of empagliflozin and dapagliflozin based on a single trial for each drug, with [redacted] results from the ITC. The ERG thus considers the results of the

			Future use of treatment is speculative. We can only reflect the care pathway used today in the submission. This is consistent with NICE guidance and committee discussion in TA398 (Section 4.18) where a specific future scenario was proposed by the manufacturer but rejected by the committee [9]. It is clear from the data presented above that prescribing of dapagliflozin is not SoC at the time of submission, there is minimal scope for displacement, and hence we do not regard this comparator of primary relevance to the decision problem. Direct economic evidence for empagliflozin vs dapagliflozin is not informative. More important is patient and prescriber choice. As the key clinical efficacy outcomes for empagliflozin and dapagliflozin are comparable, the cost-effectiveness of SGLT2i vs SoC and is the most relevant economic evidence to consider, consistent with 5.1.14 of the NICE Guide to Methods 2013.[10] This is described in [CS, Section B.3.8.3].	pooled meta-analysis conducted by the company, where it is assuming a class effect for SGLT2is, should be interpreted with caution and instead prefers the use of the Bucher ITC.

Outcomes	The outcome measures to be considered include: • symptoms of heart failure • hospitalisation for heart failure • all-cause hospitalisation • mortality • CV mortality • kidney function • adverse effects of treatment • health-related quality of life	Same	Not applicable	The ERG notes that the primary outcome from EMPEROR-R is the composite of the combined risk of CV death or HHF and it does not feature in the economic model, although the individual outcomes are also provided in the CS. The ERG also notes that the KCCQ was used in EMPEROR-Reduced to capture health status and that one of its domains includes symptoms. The ERG's clinical experts reported the KCCQ is a reasonable tool for assessing symptoms, although it doesn't tend to be routinely used in clinical practice. HRQoL data was also captured using the EQ-5D-5L, although no numerical data were reported in the clinical effectiveness sections of the CS.
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. If the technology is likely to provide similar or greater health benefits at similar or lower cost than technologies recommended in published NICE technology appraisal guidance for the same	Same	Not applicable	The company's model adheres to the decision problem for the comparison of empagliflozin and SoC. However, the company conducted a cost-comparison analysis for empagliflozin vs dapagliflozin, which the ERG disagrees with. The ERG recommends that the company uses the results from the Bucher analysis to conduct a cost utility analysis for these drugs.

	indication, a cost-comparison may be carried out. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective. The availability of any commercial arrangements for the intervention, comparator and subsequent treatment technologies will be taken into account. The cost of background therapies, such as diuretics for people with oedema, should also be included in cost effectiveness analyses.			
Subgroups to be considered	None	No subgroups were considered separately in the economic analysis	Not applicable	In response to clarification questions, the company provided the results of the Europe geographical region subgroup along with more detailed subgroups by age including a ≥75 years subgroup analysis which was recommended by the ERG's clinical experts as being potentially more representative of patients with symptomatic chronic heart failure with reduced ejection fraction in the UK.

				The ERG considers the subgroup analyses of patients by age and geographic region in EMPEROR-R suggest there may be differences in the treatment effect with empagliflozin (Section 3.3.10). In particular, the ERG is concerned about the impact the [redacted] years and Europe region subgroups are having on the overall ITT results. However, the subgroups were not powered to detect difference in treatment effect and so it is difficult to draw any firm conclusions.
Special considerations, including issues related to equity or equality	Guidance will only be issued in accordance with the marketing authorisation. Where the wording of the therapeutic indication does not include specific treatment combinations, guidance will be issued only in the context of the evidence that has underpinned the marketing authorisation granted by the regulator.	Broad prescribing of SGLT2is in primary and secondary care could reduce the inequality in access to heart failure care in the UK	Broad prescribing of SGLT2i across primary and secondary care can support the reduction in disparity in access to HF care across socio-economic groups within the UK. Only permitting a cardiologist to initiate a SGLT2i would likely widen the gap in health inequalities and lead to a delay in prescribing due to the limited resource in secondary care.[a]	Based on advice from clinical experts and the NICE recommendation for dapagliflozin in TA679, the ERG considers it unlikely that treatment with empagliflozin would be initiated without specialist input.

[a] For further details on the company's rationale please see the Company Submission, Table 1.

Abbreviations: ERG, Evidence Review Group; ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blockers; SoC, standard of care; ACEi, angiotensin-converting enzyme inhibitor; ARNi, angiotensin receptor-neprilysin inhibitor; NICE, National Institute for Health and Care Excellence; MQT, market quarter; HF, heart failure; HFrEF, heart failure with reduced ejection fraction; T2DM, Type 2 diabetes mellitus; ITC, indirect treatment comparison; SGLT2i, Sodium-glucose co-transporter-2 inhibitor; CV, cardiovascular; HHF, hospitalisation for heart failure; KCCQ, Kansas City Cardiomyopathy Questionnaire; HRQoL, health-related quality of life; EQ-5D-5L, 5-level EuroQol- 5 dimension

2.3.1 Population

The population in the key trial for empagliflozin, EMPEROR-R comprised patients aged ≥ 18 years of age with chronic HF with reduced EF defined as LVEF ≤ 40% diagnosed for at least 3 months before screening, and currently in NYHA HF class II-IV. The population is consistent with the population specified in the NICE final scope,[2] although some of the baseline characteristics of the trial population are not representative of the symptomatic HFrEF patients in the UK likely to be eligible for empagliflozin.

The key differences between EMPEROR-R and UK patients in clinical practice are that in EMPEROR-R there is a:

- Higher proportion of males;
- Lower mean age;
- Higher baseline median level of N-terminal prohormone B-type natriuretic peptide (NT-pro-BNP); and
- Lower mean LVEF.

These differences are discussed in more detail in Section 3.2.1.2 but for age and sex the ERG's clinical experts reported the differences are consistent with that seen in other clinical trials and is partly related to the strict exclusion criteria seen in clinical trials; for example, with regards to co-morbidities. In relation to the markers of disease severity, i.e. LVEF and N-terminal pro hormone B-type natriuretic peptide (NT-pro-BNP), the EMPEROR-R trial inclusion criteria was specifically targeted to focus on the recruitment of patients with more severe HFrEF and thus increased risk of outcome events. It is reported in the CS that the intent of the trial was to recruit HFrEF patients whose expected event rate for the combined risk of cardiovascular (CV) death and hospitalisation for heart failure (HHF) was at least 15% per year.

2.3.2 Intervention

Empagliflozin (Jardiance®) is an oral selective inhibitor of SGLT2[11] and is administered as a fixed 10 mg oral dose once daily in symptomatic HF. Empagliflozin currently holds European medicines agency (EMA) marketing authorisation and is recommended by NICE for the treatment of type 2 diabetes mellitus as a monotherapy (25 May 2016) or as a combination therapy with insulin or other antidiabetic drugs (25 March 2015).[12-14] In addition, the Committee for Medicinal Products for

Human Use (CHMP) adopted a positive opinion on 20 May 2021 recommending a change to the terms of the marketing authorisation for the medicinal product Jardiance. The new additional indication is the indication of relevance to this technology appraisal and was approved by the European Medicines Agency on 17 June 2021. The indication wording is, "*Empagliflozin is indicated in adults for the treatment of symptomatic chronic heart failure with reduced ejection fraction*".

The company also reported that a submission had been made to the Medicines and Healthcare products Regulatory Agency (MHRA), via the reliance route, on [redacted] and UK MHRA Marketing Authorisation for empagliflozin use in HFrEF is expected in the week commencing [redacted]. The ERG was unable to find an update on the MHRA website as of [redacted] but the company reports that on 30 July 2021, the MHRA approved empagliflozin for the indication of relevance to this technology appraisal. The indication wording is consistent with that of the European Medicines Agency.

The ERG notes that the proposed positioning of empagliflozin in UK clinical practice is as add-on therapy to SoC and this is how it was utilised in EMPEROR-R, the key source of clinical effectiveness data for empagliflozin in the CS. The inclusion criteria for EMPEROR-R required patients to be on an appropriate dose of medical therapy for HF (such as ACEi, ARB, BB, oral diuretics, MRA, Angiotensin receptor-neprilysin inhibitor [ARNi], ivabradine) consistent with prevailing local and international CV guidelines. Additionally, the drug dosage was required to be stable for at least 1 week prior to Visit 1 and during the screening period until Visit 2 (Randomisation), with the exception of diuretics which were only required to be stable for one week prior to Visit 2 (to control symptoms). The ERG notes that it is reported in the clinical study report (CSR) that [redacted] of patients were diagnosed with HF in ≤ 1 year from baseline, and it is unclear how long patients had been on stable treatment before randomisation. In response to clarification questions, the company reported that [redacted] of patients in the placebo and empagliflozin trial arms (respectively) were receiving their best tolerated treatment of guideline recommended HF therapies at baseline (clarification questions [CQ] response, Table1). However, the company reported that no data were collected on the differences between the target dose and the best tolerated dose.

The ERG notes that EMPEROR-R was a multi-centre international randomised controlled trial and so local guidelines likely varied widely. However, the ERG's clinical experts reported that the baseline SoC drugs utilised by patients in EMPEROR-R were broadly consistent with those used in the UK. The main differences flagged by the experts were that the approximately 70% ACEi or ARB usage was

potentially slightly lower than the expected 80% usage in UK clinical practice and the ARNi (sacubitril valsartan) use of approximately 20% was slightly higher than expected. This is discussed further in Section 3.2 and baseline background HF medication from EMPEROR-R is summarised in Appendix 9.2.

3 Clinical effectiveness

3.1 Critique of the methods review

The company conducted a single systematic literature review (SLR) to identify randomised controlled trial (RCT) evidence on the efficacy and safety of empagliflozin and relevant comparators in patients with chronic heart failure (HF) with reduced left ventricular ejection fraction (LVEF).

Interventions and comparators specified in the inclusion criteria for the SLR encompassed those listed as relevant to the decision problem as set out in the final scope issued by the National Institute for Health and Care Excellence (NICE).[2] Full text publications of 2214 records retrieved from the SLR were assessed for eligibility. Of the 2214 publications, 356 records representing 45 unique studies met the pre-specified inclusion criteria.

Of the 45 unique studies, three studies reported outcomes with empagliflozin. Only one of the three studies was deemed suitable for final inclusion, as the other two did not report outcomes of relevance to the economic model. The included study, EMPEROR-Reduced[15] (hereafter referred to as EMPEROR-R) was a phase III RCT comparing empagliflozin with placebo, both administered in addition to standard of care (SoC) which could include medical therapy with an angiotensin-converting enzyme inhibitor (ACEi) or angiotensin receptor-blocker [ARB], a beta-blocker (BB), and mineralocorticoid receptor antagonists (MRA). EMPEROR-R was used by the company as the primary source of clinical evidence for empagliflozin and SoC in the economic model.

As dapagliflozin was included as a comparator in the final scope,[2] analyses of the relative efficacy of empagliflozin and dapagliflozin for heart failure with reduced ejection fraction (HFrEF) was searched for by the company. However, in the absence of a head-to-head trial, studies of dapagliflozin versus alternative comparators were sought to enable an indirect comparison with empagliflozin.

Four studies describing efficacy outcomes with dapagliflozin were identified by the company. Out of the four studies, DAPA-HF[16] was selected by the company to serve as the primary study for dapagliflozin versus placebo as an add-on to SoC in comparison to empagliflozin versus placebo as an add-on to SoC in a Bucher indirect treatment comparison (ITC, discussed further in Section 3.4). The other three studies were excluded for reasons including that the population did not fully align with the National Institute for Health and Care Excellence (NICE) final scope and the outcome follow-up was too short to include in an ITC with EMPEROR-R.

Full methods and results of the SLR are reported in Appendix D of the company submission (CS). A summary of the methods, together with the evidence review group's (ERG) critique of the appropriateness of the methods adopted, is presented in Table 19. In brief, the ERG considers the methods applied by the company to be robust and likely to have identified all clinical evidence of relevance to the decision problem.

Table 19. Summary of ERG's critique of the methods implemented by the company to identify evidence relevant to the decision problem

Systematic review step	Section of CS in which methods are reported	ERG assessment of robustness of methods
Data sources	Appendix D, section D1.1.	**The ERG considers the sources and dates searched to be appropriate.** Databases searched: Medline, Embase, PsycINFO, the Cochrane Library (CENTRAL and CDSR). Additional sources: Checking reference lists of identified systematic reviews (published in 2018 to 2020), and hand-searching of conference proceedings (published in 2018 to 2020). Latest search update: October 2020.
Search strategies	Appendix D, sections D1.1.1, D1.1.2, and D1.1.3	**The ERG is satisfied that searches have identified all evidence relevant to the decision problem.** Search strategies for the literature review combined comprehensive terms for the population, interventions and study designs, using free-text and medical subject headings.
Inclusion criteria	Appendix D, sections	**The ERG considers it likely that no relevant evidence was excluded based on the eligibility criteria used.** Inclusion criteria were in line with the NICE final scope. Full reference details are available in the CS Appendix for included studies, as well as for studies excluded at full-text appraisal. Limited to English-language publications.
Screening and data extraction	Appendix D, sections D1.1.4, D1.1.5, and D1.1.7	**The ERG considers the methods for screening and data extraction to be robust.** Two reviewers independently screened titles and abstracts, and subsequently studies selected for full text appraisal, against predefined criteria, with a third reviewer consulted when consensus could not be reached. Results of the literature screening processes were summarised in PRISMA diagrams. Data extraction was carried out by one reviewer, with a second researcher independently validating and auditing extracted data.
Tool for quality assessment of included study or studies	B.2.5 & Appendix D, sections D1.3	**The ERG agrees with the company's choice of quality assessment tool.** See Appendix 9.1 for ERG validation of the quality assessment of EMPEROR-Reduced.

Abbreviations: CENTRAL, Cochrane Central Register of Controlled Trials; CDSR, Cochrane Database of Systematic Reviews; CS, company submission; ERG, Evidence Review Group; NICE, National Institute for Health and Care Excellence; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

3.2 Critique of trials of the technology of interest

In subsequent sections, the ERG focuses on aspects of trial design, conduct and external validity of EMPEROR-R, the main study that is of importance to this Single Technology Appraisal (STA). The ERG's assessment of the design, conduct and internal validity of EMPEROR-R is summarised in Table 20. The ERG agrees with the company's assessment of EMPEROR-R as being at overall low risk of bias for analysis of the co-primary outcomes, based on the full trial population (Appendix 9.1).

Table 20. Summary of the design and conduct of EMPEROR-R[15], the trial evaluating the technology of interest to the decision problem

Aspect of trial design or conduct	Section of CS providing details on trial characteristic	Summary of EMPEROR-R
Trial conduct		
Randomisation	B.2.3 (page 54 of CS)	**Appropriate** Randomised design with parallel assignment of participants in 1:1 ratio to empagliflozin, 10 mg PO once daily in addition to SoC or Placebo PO once daily in addition to the SoC. Randomisation was performed using a permuted block design with a computer pseudo-random number generator. Randomisation stratified by geographical region and history of diabetes.
Concealment of treatment allocation	B.2.3 (page 54 of CS)	**Appropriate** Treatment allocation concealed through use of IVRS/IWRS at randomisation.
Eligibility criteria	B.2.3 (page 55 of CS)	**Appropriate** Adult patients with HFrEF (LVEF ≤ 40%) diagnosed at least 3 months before screening and in the functional NYHA class II-IV. Full details are reported in B2.3.1.2
Baseline characteristics	B.2.3 (page 63 of CS)	Baseline characteristics were well balanced between the empagliflozin 10 mg and placebo groups. Full baseline characteristics from EMPEROR-R are available in Appendix 1.1.
Masking appropriate	B.2.3 (page 54 of CS)	Double-blind study. An Endpoint Adjudication Committee evaluated all reported and potential clinical events in a manner blinded to the treatment assignment.

No difference between groups in treatments given, other than intervention versus control	B.2.3, 2.6 and CQ A1. (pages 52 and 97 of CS, page 9 of CQ)	No evidence to suggest a difference between groups in treatments given additional to allocated intervention. Study drug (empagliflozin or placebo) was given in addition to SoC. The baseline SoC drug classes appeared comparable between trial arms. xxxxx xxx xxxxx of patients in the placebo and empagliflozin trial arms (respectively) were receiving their best tolerated treatment of guideline recommended HF therapies at baseline, although the ERG is unclear what proportion of patients were on the recommended target dose of the HF therapies.
Dropouts (high dropout and any unexpected imbalance between groups)	B.2.4 (page 69 of CS, and further detail in CSR - Table 10.1.2)	Proportion of patients who discontinued study treatment was low and well balanced between the two treatment groups. Treatment discontinuation was classified by either a non-fatal adverse event, request by patient or other reasons. In EMPEROR-R a total of [redacted] out of the [redacted] patients assigned to empagliflozin discontinued treatment [redacted] whereas [redacted] patients out of the [redacted] patients assigned to placebo discontinued treatment [redacted].
Outcomes assessed	B.2.3 (page 52 of CS)	All clinically relevant outcomes reported. No evidence to suggest that additional outcomes of relevance were assessed and not reported. The primary outcomes were: • Combined risk of CV death or HHF; • Total HHF (first and recurrent); • Rate of renal function decline. The key secondary outcomes were: • Risk of composite renal endpoint (chronic dialysis, renal transplant or renal insufficiency) • Risk of first HHF • Risk of CV death • Risk of death • Risk of diabetes mellitus • Change in KCCQ clinical summary score • Risk of all-cause hospitalisation Further endpoints of risk of atrial fibrillation, risk of myocardial infarction, risk of stroke, and safety were also recorded.
ITT analysis carried out	B.2.3 (page 52 of CS)	**Yes** ITT population forms the basis for analyses of efficacy.
Subgroup analyses	B.2.7 (page 84 of CS)	Prespecified subgroup analyses were based on: • Diabetes • Renal function • Gender • Race • BMI (<30 kg/m2 and ≥30 kg/m2) • Age (< 65 years and ≥ 65 years) • History of HHF in the last 12 months • Cause of HF • NYHA at baseline

		• Heart failure physiology • Baseline use of mineralocorticoid receptor antagonist • Baseline use of angiotensin receptor-neprilysin inhibitors • Geographic region • Baseline eGFR The ERG notes that geographical region, history of diabetes and eGFR at randomisation were stratification factors in EMPEROR-R. The ERG requested an additional post-hoc exploratory subgroup analysis for age with a 75-year cut-off and outcome data for the Europe geographical region subgroup.
Statistical analysis plan		
Sample size	B.2.4 (page 67 of CS)	Based on sample size calculations to detect a difference between empagliflozin 10 mg and placebo groups in the primary outcome, the company reported that 2850 patients needed to be randomised to receive empagliflozin or placebo in 1:1 manner.
Power	B.2.4 (page 67 of CS)	The company reported that, with a sample size of 2850, the study would have 90% power to detect a 20% difference in risk of a primary endpoint event for a two-sided test with α=0.05.
Analysis sets		• Randomised set (RS): All randomised patients, whether treated or not. • Treated set (TS): All patients who were dispensed study medication and were documented to have taken at least one dose of investigational treatment. • Treated Set-Follow-up (TS-FU): All patients in the TS for whom a follow-up visit was performed (i.e. values of planned assessments: KCCQ, EQ-5D, vital signs or lab data reported) between 23 and 45 days after last intake of study medication. The TS-FU did not include patients with only telephone FU visits and for whom no planned measurements were taken.
Analysis for estimate of effect	B.2.4 (page 67 of CS)	**Primary endpoints** The company reports that for each of the primary endpoints, superiority of empagliflozin over placebo was evaluated with a two-sided test. The overall type I error rate for the trial was preserved at α=0.05. Due to the amount of α spent on the interim analysis, the remaining two-sided α level for the final analysis was 0.0496. The primary analysis was a Cox PH regression with factors treatment, geographical region, diabetes status at baseline, age, gender, LVEF, and baseline eGFR. Following the ITT principle, the primary analysis was based on RS using all data up to the end of the planned treatment period (i.e., excluding events and time at risk after the protocol-specified treatment discontinuation for patients who completed the treatment period but including the data after end of treatment for patients not completing the treatment phase as planned). Patients without a specific endpoint event were censored at the last date the patient was known to be event free or at the end of the planned treatment period, whichever was earlier. When violation of the PH assumption was observed, groups of patients for which the proportionality assumption held were identified, and a stratified Cox regression was performed. **Secondary endpoints** Occurrence of adjudicated HHF was analysed by a joint frailty model that accounted for the dependence between recurrent HHF and CV death. The

		primary analysis included all data until completion of the planned treatment phase, including the data after end of treatment for patients not completing the treatment phase as planned. The model included the same covariates used for the analysis of the primary endpoint. Slope in change from baseline of eGFR was analysed by a random coefficient model allowing for random intercept and random slope per patient, with the same factors used for the primary endpoint and the additional factors time, treatment-by-time and baseline eGFR interaction as linear covariates. The model included all on-treatment change from baseline. This endpoint was tested with a two-sided α of 0.001.

Abbreviations: ERG, Evidence Review Group; CS, Company submission; PO, per os [oral administration]; SoC, Standard of care; IVRS/IWRS, interactive voice response system/integrated web response system; HFrEF, heart failure with reduced ejection fraction; LVEF, left ventricular ejection fraction; NYHA, New York Heart Association; CQ, clarification question; HF, heart failure; CV, cardiovascular; HHF, hospitalisation for heart failure; KCCQ, Kansas City Cardiomyopathy Questionnaire; ITT, intention to treat; BMI, body mass index; eGFR, estimated Glomerular Filtration Rate; RS, Randomised set; TS, Treated set; TS-FU, Treated Set-Follow-up; EQ-5D, EuroQol- 5 dimension; FU, follow-up.

3.2.1 External validity of EMPEROR-R

The ERG considers that while the findings of EMPEROR-R can be broadly applied to UK practice, the trial population represents a 'high-risk' subpopulation of the patients within UK clinical practice who would typically receive empagliflozin for the proposed indication.

The trial recruited males and females aged ≥ 18 years of age with chronic HF diagnosed for at least 3 months. Patients were included if they were identified to have moderate to severe HFrEF (LVEF ≤40%, New York heart association [NYHA] II-IV). However, the ERG's clinical experts highlighted that the participants contributing data to the trial were less well than a typical UK HF patient, as indicated by the low ejection fraction and high NT pro-BNP. The ERG considers that the trial population is focused on sicker patients than is generally seen in UK practice and therefore the patients in EMPEROR-R are likely to have increased risk of poorer health outcomes than expected in clinical practice.

The ERG's clinical experts also noted that the proportion of participants in EMPEROR-R who were female (~24%) was smaller than would be expected in a typical HF population (30-40%), and that the trial population (mean age ~67 years) may be younger than those that would be seen in clinical practice (typically ~10 years older than those seen in EMPEROR-R). However, during clarification the company provided additional subgroup analyses by patient age at <65; 65–<75; and ≥75 years (see Section 3.3.10 for further discussion of the subgroups).

The study compared intervention with empagliflozin to placebo, both of which were administered in addition to SoC which could include medical therapy with ACEi/ARB or ARNi, BBs and MRAs. The

ERG's clinical advisors considered this to be reflective of SoC in the UK, but did note that the proportion of participants on an ACEi or ARB in the EMPEROR-R trial (~70%) was lower than would be expected in a typical HF population in the UK (~80%). The ERG's clinical experts also noted that the proportion of patients in EMPEROR-R with an ARNi (~19%) was slightly higher than would be expected in a UK population, and potentially reflective of a population whose HF is more severe, capturing patients who are still symptomatic. Patients were also required to be on an appropriate dose of medical therapy for HF consistent with local and international guidelines, stable for at least one week prior to screening and during the screening period until randomisation (between 4 to 28 days). As discussed in Section 2.3.2, the appropriate dose was not necessarily the target dose and the proportion of patients who were on the target dose at baseline in EMPEROR-R is unclear.

The trial informing the CS was a multicentre international study, conducted across sites in North America (11%), Latin America (34%) Europe (36%), Asia (13%) and other regions (5%). The company provided outcome data for the pooled ITT population without subgroup analysis by geographic region, arguing that this total population is representative of UK population. As such, with the suggestion that this population is more generalisable to the ethnically diverse UK population, the company utilised this population data for the economic analysis. At the ERG's request, during clarification the company provided additional subgroup results for participants from the Europe geographical region for each arm of EMPEROR-R. The results could suggest a reduced efficacy in the primary composite outcome within the European subgroup (hazard ratio [HR] 0.94; 95% confidence interval [CI]: 0.74 to 1.18) compared to the ITT population (HR 0.75; 95% CI: 0.65 to 0.86), as well as total hospitalisation for heart failure (HHF) (HR 0.96; 95% CI: 0.70 to 1.33 vs 0.70; 95% CI: 0.58 to 0.85, respectively [discussed further in Section 3.3.10.2]). However, the ERG's clinical experts advised that it would be unlikely that any differences in outcomes with empagliflozin would be dependent on geographical region or race.

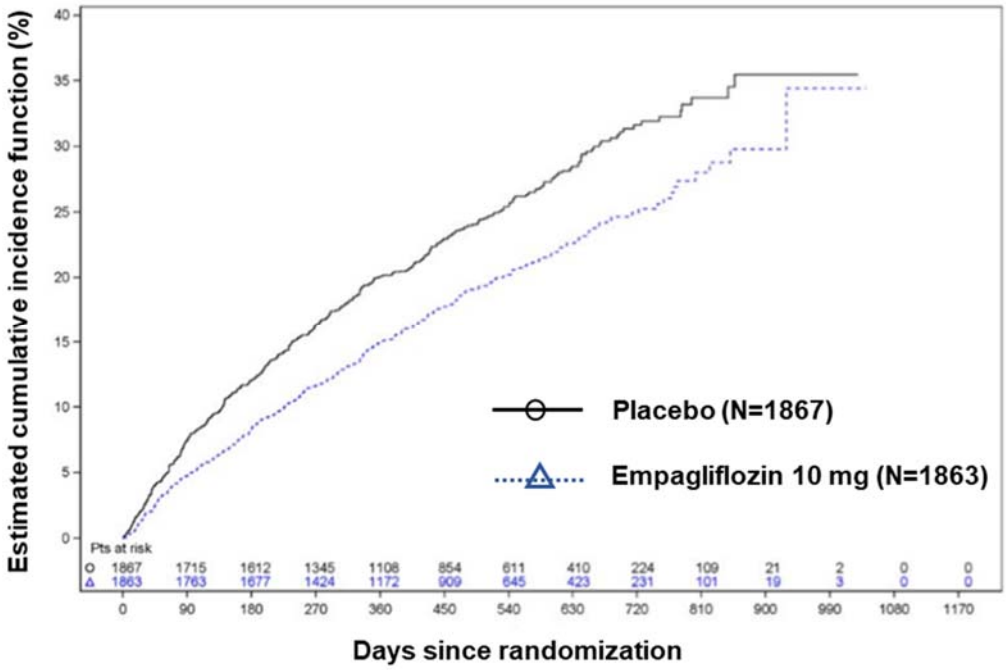
3.3 Clinical effectiveness results from EMPEROR-Reduced

3.3.1 Combined risk of cardiovascular death or HHF

There was a median follow-up of 16 months for the primary composite outcome of cardiovascular (CV) death or HHF in EMPEROR-R. The results demonstrated that fewer patients in the empagliflozin group (361/1863 patients, 19.4%) experienced an event compared to in the placebo group (462/1867 patients, 24.7%). The company presented the estimated cumulative incidence plots of CV death or first HHF, considering non-CV death as a competing risk (Figure 2). The ERG notes from the

plots for the individual outcomes of CV death and HHF that the benefit for empagliflozin appears to be driven by HHF.

Figure 2. Estimated cumulative incidence function for time to the first event of adjudicated CV death or HHF in all randomised patients (Reproduced from CS, Figure 7)



Source: EMPEROR-Reduced CSR, Figure 11.1.1.1:1 [17]

The company conducted a Cox regression of the data for all randomised patients adjusted for age, baseline estimated glomerular filtration rate (estimated glomerular filtration rate [eGFR]; based on the serum creatinine Chronic Kidney Disease Epidemiology Collaboration equation [(CKD-EPI)cr][18]), region, gender, treatment, baseline diabetes status and LVEF which showed a statistically significant reduction in the risk of CV death or HHF with empagliflozin compared with placebo (HR 0.75; 95% CI 0.65 to 0.86, p<0.0001). In addition, the company conducted three sensitivity analyses of the primary endpoint:

1. Multiple imputation analysis to account for missing follow-up data in 42 patients;
2. Removing the covariate adjustments applied in the primary analysis;
3. Adjusting for non-CV death as a competing risk.

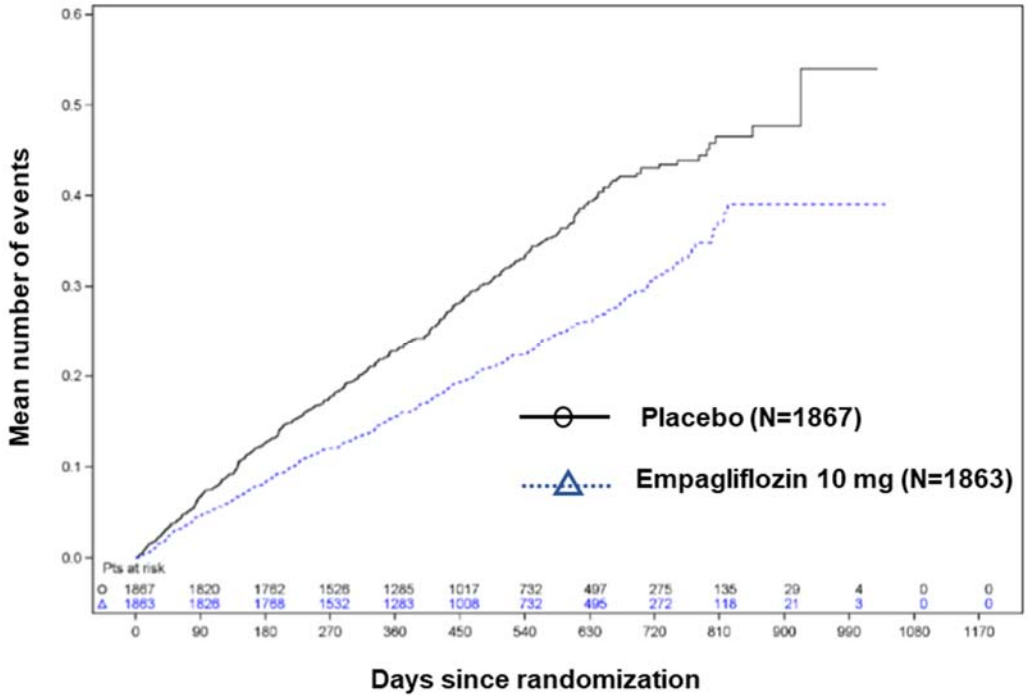
The ERG notes that the results from the sensitivity analyses were all consistent with the results of the primary analysis (CS, Table 18).

3.3.2 Hospitalisation for heart failure

3.3.2.1 Total number of hospitalisations for heart failure (first and recurrent)

The total number of hospitalisations for HF was lower in the empagliflozin group than in the placebo group with 388 events and 553 events, respectively. Figure 3 shows the mean cumulative incidence plot of HHF over time, with the benefit from empagliflozin occurring soon after randomisation and being maintained throughout the trial.

Figure 3. Mean cumulative function for occurrence of adjudicated HHF (first and recurrent) in the randomised set (Reproduced from CS, Figure 8)



Source: EMPEROR-Reduced CSR, Figure 11.1.2.1.1:1 [17]

The primary analysis of total hospitalisations was conducted using a joint frailty model[19] with CV death as a competing risk and showed a statistically significant reduction in the risk of recurrent HHF with empagliflozin compared to placebo (HR 0.70; 95%CI: 0.58 to 0.85, p<0.001). The ERG notes that the use of the joint frailty model is a method of addressing unobserved heterogeneity in the underlying hazard and that the company also conducted sensitivity analyses that explored the use of alternative methods for the occurrence of adjudicated HHF which are discussed below. The company reported that the hazard of recurrent HHF, where CV death was included as a competing risk, was positively correlated with CV death because the frailty exponent was greater than zero.

The company conducted sensitivity analyses for the occurrence of adjudicated HHF (first and recurrent) which included the removal of covariates, and including all-cause mortality instead of CV death as a competing risk in the joint frailty model. The results of the sensitivity analyses were consistent with the results of the primary analysis (CS, Table 19).

3.3.2.2 First adjudicated hospitalisation for heart failure

The estimated cumulative incidence of first adjudicated HHF curves, considering all-cause mortality as a competing risk (Figure 4), showed [redacted] Figure 3). The risk of adjudicated HHF was [redacted] with empagliflozin treatment compared with placebo ([redacted] when calculated using the Cox regression model adjusted for age, baseline eGFR (CKD-EPI)cr, region, baseline diabetes status, gender, baseline LVEF, and treatment.

Figure 4. Estimated cumulative incidence function for time to the first adjudicated HHF with all-cause mortality as a competing risk, randomised set (Reproduced from CS, Figure 11)

Source: EMPEROR-Reduced CSR, Figure 11.1.2.3.1:1 [17]

3.3.3 First and recurrent all-cause hospitalisation

All-cause hospitalisation occurred in [redacted] patients in the empagliflozin group [redacted] compared with in the placebo group [redacted]. In addition, the total number of hospitalisation events (first and recurrent all-cause hospitalisation) was [redacted] in the empagliflozin group ([redacted] hospitalisation events) compared to in the placebo group ([redacted] hospitalisation events). The company used a joint frailty model to account for the dependence between recurrent all-cause

hospitalisation and all-cause mortality, and the results showed [redacted] in the risk of recurrent all-cause hospitalisation with empagliflozin treatment compared to placebo ([redacted]

The mean cumulative incidence plots of all-cause hospitalisation in the empagliflozin and placebo groups (Figure 5) [redacted] (Figure 3). The results of Cox regression demonstrated an [redacted] reduction in risk of first all-cause hospitalisation with empagliflozin compared to placebo and the difference [redacted]. The company reported in their response to clarification questions that the proportion of patients who were hospitalised for a non-HF event in EMPEROR-R was [redacted]). The ERG [redacted]

Figure 5. Mean cumulative function plot for occurrence of all-cause hospitalisation (first and recurrent), randomised set (Reproduced from CS, Figure 15)

Source: EMPEROR-Reduced CSR, Figure 11.1.2.5.1:1 [17]

3.3.4 Renal function

3.3.4.1 Deterioration of renal function

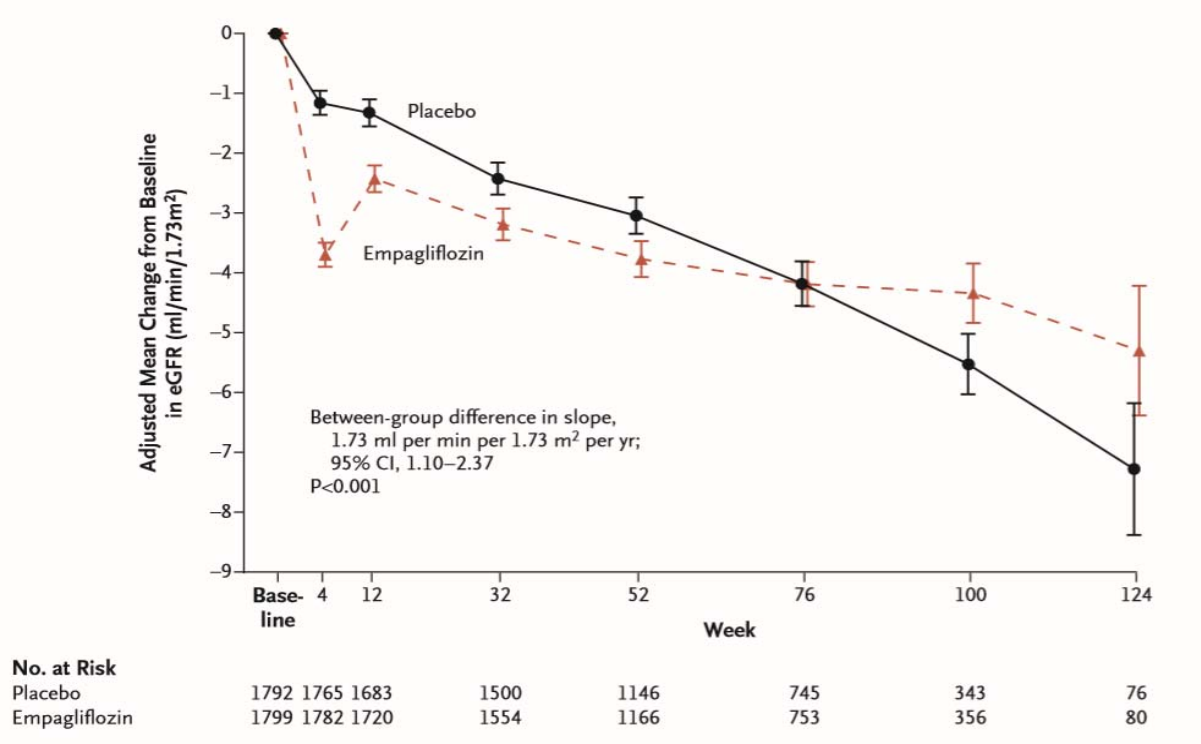
Estimation of GFR for the analyses of renal function were based on the serum creatinine (CKD-EPI)cr.[18] The primary analysis of mean slope of change in eGFR [mL/min/1.73 m^2] from baseline used the treated set rather than the intention-to-treat (ITT) population (randomised set) and included measurements up to one day after the last intake of study medication. The estimated slope was -0.55 ± 0.23 mL/min/1.73 m^2 per year with empagliflozin, and -2.278 ± 0.23 mL/min/1.73 m^2 per year with

placebo. The estimated between-group difference in mean slope was 1.73 mL/min/1.73 m^2 per year, indicating a lower rate of decline with empagliflozin compared with placebo (95% CI: 1.10 to 2.37, p<0.001 [Figure 6]).

The ERG notes that the initial dip in eGFR seen at the start of the empagliflozin curve in Figure 6 is attributed by the company to, *"a reversible functional change in intrarenal haemodynamics commonly observed with Sodium-glucose co-transporter-2 inhibitor (SGLT2is) and is not associated with an excess risk of investigator-reported acute kidney injury[20]"*. The company provided further detail to support this rationale in their response to clarification question A4b. However, the ERG notes that up until week 76, empagliflozin is associated with a greater reduction in eGFR compared with placebo and beyond week 76 there is heavy censoring of patients in both treatment groups. Additionally, in response to a clarification question, the company provided numerical data to accompany Figure 6 (clarification question [CQ] response Table 5) and the data do not demonstrate a statistically significant difference between empagliflozin and placebo at weeks 100 and 124.

The company also reported the adjusted mean eGFR change from baseline to follow-up, although this is in the randomised set and the ERG is unclear of the adjustments applied. The results of this analysis showed the adjusted mean eGFR change from baseline was 3.3 mL/min/1.73m^2 (95% CI: 1.8 to 4.8) for empagliflozin versus placebo. The ERG notes that this is lower than the 5 or greater mL/min/1.73m^2 minimal clinically meaningful difference observed in other studies (e.g. Mayne *et al.*)[21] but the ERG's clinical experts consider that it is likely that there will be a long-term renal benefit with empagliflozin.

Figure 6. Changes in the estimated glomerular filtration rate, based on the treated set and measurements up to one day after the last intake of study medication (Reproduced from CS, Figure 9)



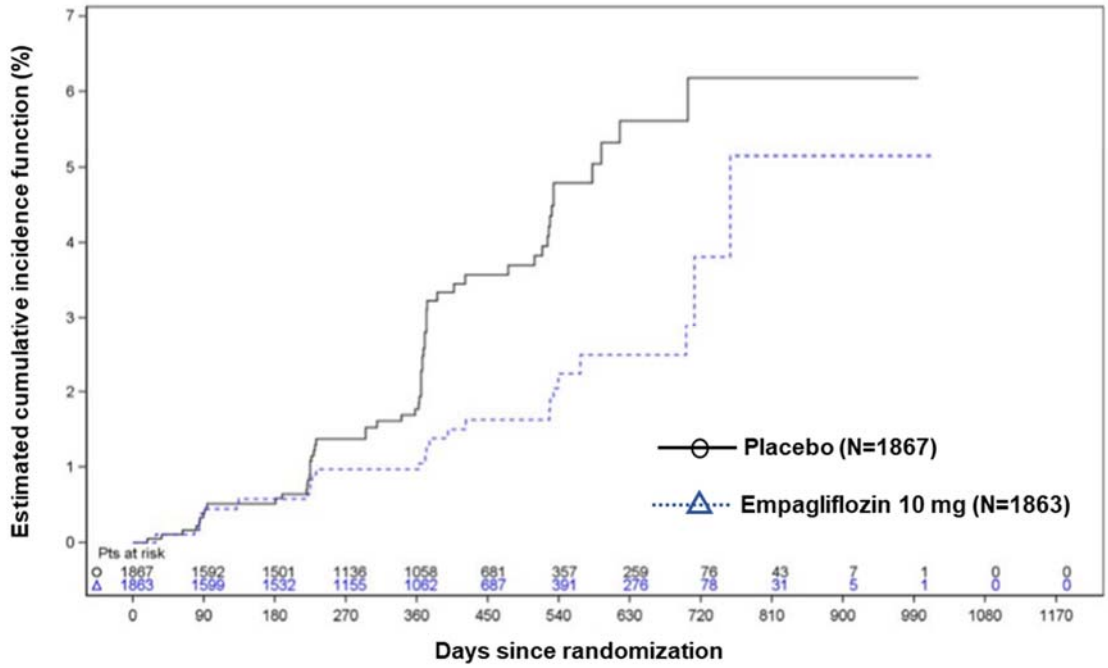
Source: Packer et al 2020 [7]

Note: Graph shows the adjusted mean changes from baseline in the eGFR as calculated using the Chronic Kidney Disease Epidemiology Collaboration equation. The bars indicate the standard error. The on-treatment data were analysed with a mixed model for repeated measures (MMRM). Age and baseline eGFR were included as linear covariates, while sex, region, baseline LVEF, baseline diabetes status, last projected visit based on dates of randomisation and trial closure, baseline eGFR according to visit, and visit according to treatment interactions were included as fixed effects.

3.3.4.2 Composite renal outcome

Time to the first event in the composite renal endpoint of chronic dialysis, renal transplant or sustained reduction in eGFR (CKD-EPI)cr, using the randomised set, is shown in Figure 7. The number of patients with a composite renal endpoint event along with a breakdown by type of event is summarised in Table 21. The ERG notes that the overall incidence of the composite renal endpoint was [redacted] ([redacted] patients [redacted%] in the empagliflozin group and [redacted] patients [redacted%] in the placebo group). However, the risk of the composite renal endpoint was markedly lower with empagliflozin compared to placebo (HR 0.50, 95%CI: 0.32 to 0.77, nominal p = 0.002). The ERG notes that [redacted]

Figure 7. Estimated cumulative incidence function plot for time to the first event of the composite renal endpoint, randomised set (Reproduced from CS, Figure 10)



Source: EMPEROR-Reduced CSR, Figure 11.1.2.6.1:1 [17]

Table 21. Cox regression analysis of time to first renal event[¶], randomised set (Reproduced from CS, Table 20)

Time to composite renal outcome*	Placebo (N=1867)	Empagliflozin (N=1863)
Patients with the composite renal endpoint, N (%)	[redacted]	[redacted]
Sustained eGFR reduction ≥40% as the first event	[redacted]	[redacted]
Sustained eGFR <15 mL/min/1.73 m^2 (baseline ≥30) or <10 mL/min/1.73 m^2 (baseline <30) as the first event	[redacted]	[redacted]
Chronic dialysis as the 1st event	[redacted]	[redacted]
Renal transplant as the 1st Event	[redacted]	[redacted]
Incidence rate per 100 years at risk	[redacted]	[redacted]
Hazard ratio *vs.* placebo (95% CI), composite renal outcome	0.50 (0.32 - 0.77)	
Nominal p-value	0.0019	

Source: EMPEROR-Reduced CSR, Table 11.1.2.6:1 [17]

Abbreviations: eGFR, estimated glomerular filtration rate; CI, confidence interval; RS, randomised set.

*The composite renal endpoint was comprised of chronic dialysis (with a frequency of twice per week or more for at least 90 days), renal transplant, sustained reduction in eGFR from baseline of ≥40%, sustained eGFR <15 mL/min/1.73m^2 for patients

with baseline eGFR ≥30 mL/min/1.73m^2, or sustained eGFR <10 mL/min/1.73m^2 for patients with baseline eGFR <30mL/min/1.73m^2. Sustained was determined by two or more consecutive post-baseline central laboratory measurements separated by at least 30 days (the first to last of the consecutive eGFR values).

¶Cox regression model included covariates age, baseline eGFR (CKD-EPI)cr, region, baseline diabetes status, sex, baseline LVEF, and treatment.

3.3.4.3 ≥ 5 or ≥10 ml/min/1.73 m^2 decline in eGFR

Following advice from clinical experts that a change in eGFR of ≥ 5 ml/min/1.73 m^2 is clinically meaningful the ERG requested the company provide data on patients with a ≥ 5ml/min reduction in eGFR and patients with a ≥ 10ml/min reduction in eGFR (Table 22). The company reported that Table 4 of the company response to CQ (Table 22), "[redacted]". However, the ERG considers the data mostly show [redacted] at each timepoint with empagliflozin compared with placebo. The ERG acknowledges that from week [redacted]. However, the ERG considers caution should be taken when drawing conclusions from these data as the adjusted mean eGFR change from baseline was 3.3 mL/min/1.73m2 (95% CI: 1.8 to 4.8) for empagliflozin versus placebo and the ERG's clinical experts consider that it is likely that there will be a long-term renal benefit with empagliflozin.

Table 22. proportion of patients who have a ≥ 5ml/min and ≥10mL/min reduction in eGFR (Reproduced from CQ response Table 4)

Analysis visit	Treatment	Proportion of patients with a ≥5mL/min reduction in eGFR, n (%)	Proportion of patients with a ≥10mL/min reduction in eGFR, n (%)
Number of patients in analysis set	Placebo	[redacted]	
	Empagliflozin (10mg)	[redacted]	
Week 4	Placebo	[redacted]	[redacted]
	Empagliflozin (10mg)	[redacted]	[redacted]
Week 12	Placebo	[redacted]	[redacted]
	Empagliflozin (10mg)	[redacted]	[redacted]
Week 32	Placebo	[redacted]	[redacted]
	Empagliflozin (10mg)	[redacted]	[redacted]
Week 52	Placebo	[redacted]	[redacted]

	Empagliflozin (10mg)	[redacted]	[redacted]
Week 76	Placebo	[redacted]	[redacted]
	Empagliflozin (10mg)	[redacted]	[redacted]
Week 100	Placebo	[redacted]	[redacted]
	Empagliflozin (10mg)	[redacted]	[redacted]
Week 124	Placebo	[redacted]	[redacted]
	Empagliflozin (10mg)	[redacted]	[redacted]
Week 148	Placebo	[redacted]	[redacted]
	Empagliflozin (10mg)	[redacted]	[redacted]

Source: CTR, 1245_121 – Randomised set, observed case – after discontinuation of study medication

3.3.4.4 Acute dialysis

The ERG noted that the composite renal outcome included only the first renal events that were chronic dialysis events and not acute dialysis events or chronic dialysis events that occurred after another renal composite event. The ERG, therefore, requested details on acute renal dialysis in EMPEROR-R during the clarification stage. The company reported in their clarification response (CQ A4d) that while data on the need for dialysis after baseline was collected in EMPEROR-R, it wasn't specified whether it was for acute or chronic dialysis. However, as part of the response to CQ A4b, the company also reported end-stage renal disease events and the number of patients requiring chronic dialysis, although it is unclear which analysis sets these data relate to.

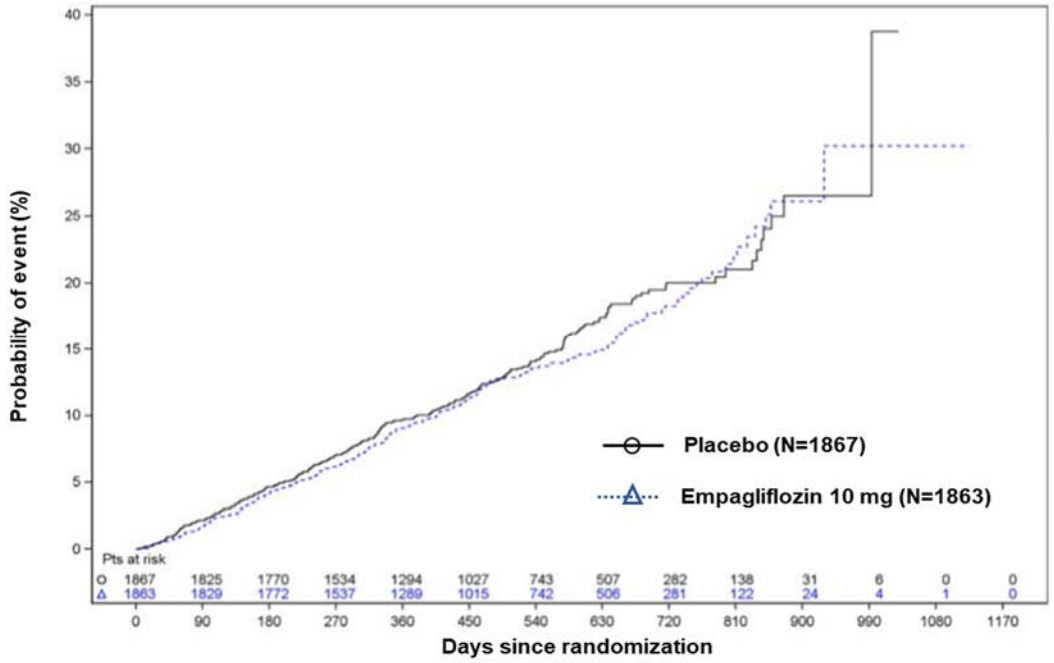
The ERG notes that [redacted] patients required dialysis (acute or chronic) in the empagliflozin arm ([redacted] patients [redacted]%]) compared to in the placebo arm ([redacted] patients [redacted]%]). [redacted] developed end-stage renal disease in the empagliflozin arm compared to [redacted] patients in the placebo arm [redacted]. Additionally, it was reported that only [redacted] required chronic dialysis in the empagliflozin arm compared to [redacted] patients in the placebo arm [redacted] These data support the company view that empagliflozin offers some degree of renal protection compared to SoC.

3.3.5 All-cause mortality

Death from any cause occurred in 249 patients (13.4%) in the empagliflozin group of EMPEROR-R and 266 patients (14.2%) in the placebo group. The Kaplan-Meier estimate of time to all-cause mortality in the randomised set show the curves for empagliflozin and placebo overlap at multiple timepoints (Figure 8).

The company's Cox regression of time to all-cause mortality data for all randomised patients demonstrated no statistical difference in the risk of death from any cause with empagliflozin compared with placebo (HR 0.92, 95% CI: 0.77 to 1.10; p=0.35).

Figure 8. Time to all-cause mortality, Kaplan-Meier estimate in randomised set (Reproduced from CS, Figure 12)



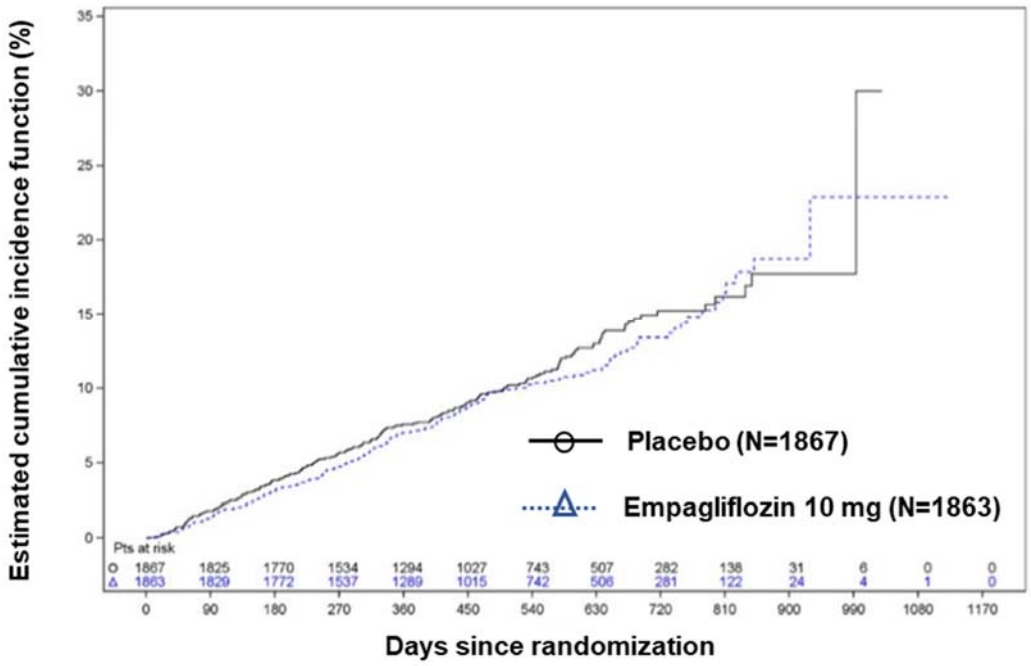
Source: EMPEROR-Reduced CSR, Figure 11.1.2.4.1:1 [17]

3.3.6 Cardiovascular mortality

The company reported that the majority of the deaths recorded during EMPEROR-R were a result of CV causes, such as sudden cardiac death or HF death. A total of 187 patients (10.0%) in the empagliflozin group and 202 patients (10.8%) in the placebo group were adjudicated as having a CV death.

The estimated cumulative incidence plot of adjudicated CV death in randomised patients, considering non-CV death as a competing risk suggest similar mortality with empagliflozin compared with placebo (Figure 9). Additionally, the risk of CV death with empagliflozin compared to placebo did not reach statistical significance (HR 0.92, 95%CI: 0.75 to 1.12; p=0.41).

Figure 9. Estimated cumulative incidence function plot for time to adjudicated CV death, considering non-CV death as a competing risk, randomised set (reproduced from CS, Figure 13)



Source: EMPEROR-Reduced CSR, Figure 11.1.2.4.2:1 [17]

3.3.7 Kansas City Cardiomyopathy Questionnaire (KCCQ)

As discussed in Section 2.3, the change from baseline in health status was assessed by the Kansas City Cardiomyopathy Questionnaire clinical summary score (KCCQ-CSS) at week 52 with interim data collected at weeks 12 and 32. The clinical summary score measures HF symptom frequency, symptom burden, and physical limitations. In addition, data were collected for the KCCQ total symptom score (KCCQ-TSS) and KCCQ overall summary score (KCCQ-OSS) which incorporates quality of life.

The ERG notes that there is a disparity in the analysis set used for the KCCQ outcomes reported in the CS, Table 21 in the clinical effectiveness section compared with the data reported in the CS, Figure 20 in the cost-effectiveness section. In the CS, Figure 20, the results for KCCQ-CSS change from baseline are reported using the randomised set, whereas in the CS, Table 21 the treated set are

used. The ERG also notes from the company response to clarification that sensitivity analyses for KCCQ-CSS in the randomised set were conducted, where patients who died, had a score of 0 (worst score) imputed at all subsequent scheduled visits after the date of death. In response to clarification questions, the company reported the results for KCCQ-CSS in the randomised set with and without the imputation for death. The ERG notes that, [redacted] (Table 23).

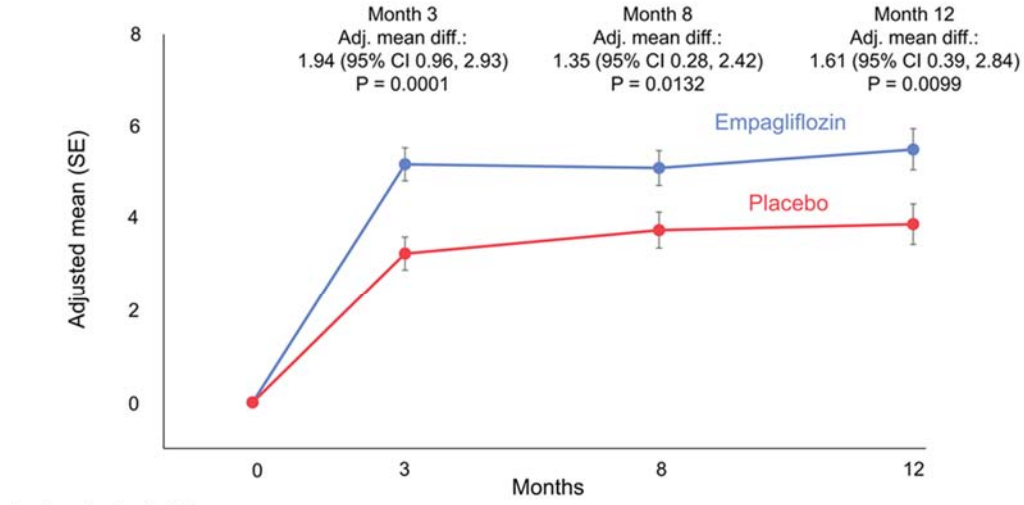
Table 23. Mean change in KCCQ-CSS from baseline to week 12, 32 and 52 (Reproduced from CQ response, Table 8)

Mean change from Baseline to:	Without imputation			With imputation		
	Placebo	Empagliflozin (10mg)	Diff	Placebo	Empagliflozin (10mg)	Diff
Baseline						
N	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mean (SE)	[redacted]	[redacted]		[redacted]	[redacted]	
Week 12a						
Adjusted Mean, SE, [95% CI], p [a]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 32 [a]						
Adjusted Mean, SE, [95% CI], p [a]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 52 [a]						
Adjusted Mean, SE, [95% CI], p [a]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source data: Table 15.2.3.6:5, RS, OC-AD without imputation for death; Table 15.2.3.6:1, RS, OC-AD with imputation;
a. adjusted for covariates variables, MMRM; both analyses used the randomised set, observed case after discontinuation.

Figure 10 (CS, Figure 20) reports the change in KCCQ-CSS over time for the treated set without imputation for death. [redacted] change from baseline KCCQ-CSS was statistically significant, favouring empagliflozin at each timepoint. The ERG notes that the largest changes in KCCQ-CSS were observed in both the empagliflozin and placebo treatment groups during the first three months of treatment and the effect then appears to plateau and generally remain stable until the end of study follow-up at week 52 (12 months). The ERG notes that the adjusted mean difference in KCCQ-CSS score was [redacted] at each timepoint and that published literature suggests a minimum of a 5 point difference in the KCCQ overall score is clinically significant.[23] The ERG acknowledges that the KCCQ-CSS is a subset of the KCCQ overall score but nevertheless considers it important to highlight that the mean difference in KCCQ-CSS scores between empagliflozin and placebo [redacted] (Table 23).

Figure 10. Effect of empagliflozin vs placebo on mean KCCQ-CSS over time (CS, Figure 20)



Analyzed patients (n)				
Placebo	1753	1732	1568	1218
Empagliflozin	1776	1755	1618	1239

Source: Butler et al 2021 [22]

In terms of the outcomes of KCCQ-TSS and KCCQ-OSS, the results were reported in the CS using the treated set and the ERG notes that in the clinical study report (CSR) the data for these outcomes [redacted] However, the ERG's preferred data set is the randomised set for both the baseline and 52 week results, and so the ERG has extracted the 52 week data for the randomised set with the imputation for death for all three outcomes (Table 24). The ERG notes that these data show baseline scores between treatment

arms [redacted] (CSS, TSS and OSS; Table 24).In addition, the results for KCCQ-TSS and KCCQ-OSS using the randomised set with imputation for death were [redacted]; Table 24).

Table 24. Summary of additional KCCQ secondary endpoints from EMPEROR-Reduced study (Adapted from CQ response, Table 2 and CS, Table 21)

Endpoint	Placebo	Empagliflozin 10 mg
Baseline KCCQ scores, mean (SE), randomised set (RS) with imputation for death[$]		
Number of analysed patients	[redacted]	[redacted]
KCCQ-CSS	[redacted]	[redacted]
KCCQ-TSS	[redacted]	[redacted]
KCCQ-OSS	[redacted]	[redacted]
QoL measured by KCCQ at 52 weeks[§], randomised set (RS) with imputation for death[$]		
Number of analysed patients	[redacted]	[redacted]
Change in clinical summary score at 52 weeks, adjusted mean (SE)	[redacted]	[redacted]
Adjusted mean change from baseline (95% CI)	[redacted]	
Nominal p-value	[redacted]	
Change in overall summary score (OSS) at 52 weeks, adjusted mean (SE)	[redacted]	[redacted]
Adjusted mean change from baseline (95% CI)	[redacted]	
Nominal p-value	[redacted]	
Change in total symptom score at 52 weeks, adjusted mean (SE)	[redacted]	[redacted]
Adjusted mean change from baseline (95% CI)	[redacted]	
Nominal p-value	[redacted]	

Source: EMPEROR-Reduced CSR[17]

Abbreviations: AE, adverse event; AF, atrial fibrillation; CI, confidence interval; ECG, electrocardiogram; HbA1c, glycated haemoglobin; HR, hazard ratio; KCCQ, Kansas City Cardiomyopathy Questionnaire; MI, myocardial infarction; QoL, quality of life; RS, randomised set, TS, treated set.

Note: KCCQ scores were analysed using mixed model for repeated measures (MMRM). [redacted]

$ For patients who died, a worst score (score of 0) is imputed at all subsequent scheduled visits.

§The clinical summary score on the Kansas City Cardiomyopathy Questionnaire ranges from 0 to 100, with higher scores indicating a better quality of life. Analysis of PRO data with a MMRM at week 52 was based on the treated set and using on-treatment values only.

[a] Data extracted from CSR by the ERG

3.3.8 EQ-5D-5L

The company reported that there were no relevant differences between the empagliflozin and placebo treatment groups with regards to health-related quality of life (HRQoL) as assessed by the EQ-5D-5L questionnaire. No numerical data were presented in the CS and so the ERG is unable to comment further on this outcome.

3.3.9 Further secondary clinical endpoints

Results of further exploratory secondary endpoints from EMPEROR-R trial, including time to onset of diabetes mellitus (DM) in patients with pre-DM were provided in the CS Section B.2.6.2 but are not of direct relevance to the NICE final scope or economic model and therefore not discussed further by the ERG.

3.3.10 Subgroup analyses

As already noted in Table 20, there were numerous pre-specified subgroup analyses for the efficacy endpoints of EMPEROR-R were conducted. The ERG acknowledges the statistical limitations in interpreting the findings of the subgroup analyses as they were not powered to detect statistically significant differences and that the company reports they were hypothesis generating.

The ERG notes that results from the subgroup analyses were mostly limited to the primary composite outcome of the combined risk of CV death or HHF and that generally the effect of empagliflozin was consistent across the pre-specified subgroups (CS, Figure 16 and CS Appendix E). The company reported that the magnitude of the benefit was smaller in NYHA class III-IV (more severe HF) subgroup versus NYHA class II (less severe HF) subgroup at baseline, although subgroup analyses of other measures of HF severity, such as LVEF and N-terminal pro hormone B-type natriuretic peptide (NT-pro-BNP) levels, did not support the same directionality of effect.

3.3.10.1 Age subgroups

The ERG requested subgroup analyses by age during the clarification stage due to advice from clinical experts that the mean age in the trial population was younger than expected in clinical practice (where the expected age is closer to 75 years or older). The company provided results for subgroups aged <65, 65–<75, and ≥75 years (Table 25). The ERG notes that for the primary composite outcome and total HHF (first and recurrent) the efficacy of empagliflozin may [redacted]

[redacted] In particular, the ERG notes that the benefit with empagliflozin compared to placebo in [redacted] The ERG considers that the primary composite outcome and HHF results for the overall population may therefore be potentially being driven by the [redacted] subgroup.

In addition, the ERG notes that there is a [redacted] mean difference in change in KCCQ-CSS score from baseline with empagliflozin compared to placebo in the [redacted] years subgroup compared to the [redacted] years subgroup and the overall population. The ERG considers the difference in KCCQ-CSS score to be of particular importance given the impact of KCCQ-CSS in the economic model (discussed further in Section 4).

Table 25. Primary and secondary outcomes in EMPEROR-Reduced by age subgroup (<65 years, 65-75 years, ≥75 years) (Adapted from CQ response, Table 13)

Subgroup category	N with event / N analysed (%)		Hazard ratio (95% CI), p
	Placebo	Empagliflozin	
Time to the first event of adjudicated CV-death or HHF			
Overall	[redacted]	[redacted]	[redacted]
<65 years	[redacted]	[redacted]	[redacted]
65–<75 years	[redacted]	[redacted]	[redacted]
≥75 years	[redacted]	[redacted]	[redacted]
CV-death[c]			
Overall	[redacted]	[redacted]	[redacted]
<65 years	[redacted]	[redacted]	[redacted]
65–<75 years	[redacted]	[redacted]	[redacted]
≥75 years	[redacted]	[redacted]	[redacted]
All-cause mortality			
Overall	[redacted]	[redacted]	[redacted]
<65 years	[redacted]	[redacted]	[redacted]
65–<75 years	[redacted]	[redacted]	[redacted]
≥75 years	[redacted]	[redacted]	[redacted]
Total HHF (first and recurrent)[b]			
	Total HHF / N analysed		Hazard ratio (95% CI), p
Overall	[redacted]	[redacted]	[redacted]
<65 years	[redacted]	[redacted]	[redacted]

65–<75 years	[redacted]	[redacted]	[redacted]
≥75 years	[redacted]	[redacted]	[redacted]
KCCQ-CSS change from baseline at week 52[d]			
	Adjusted mean (SE)		Mean difference (95% CI), p
Overall	[redacted]	[redacted]	[redacted]
<65 years	[redacted]	[redacted]	[redacted]
65–<75 years	[redacted]	[redacted]	[redacted]
≥75 years	[redacted]	[redacted]	[redacted]

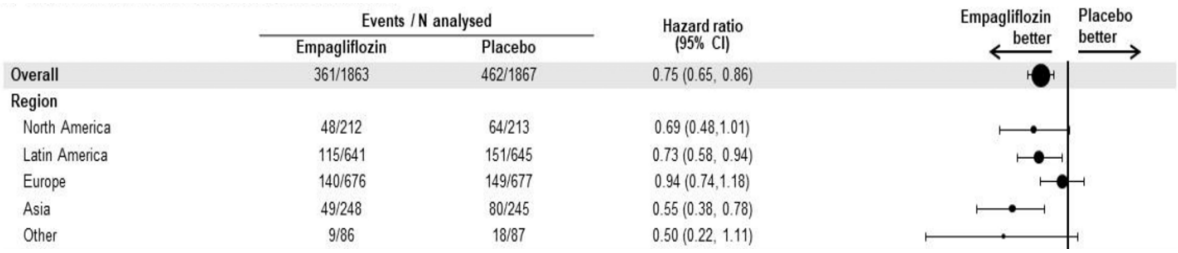
Source: CTR, Appendix 2, 18.1 [Pg 2-24] (all other outcomes); Appendix 2, Table 37.1.1 (KCCQ).

[a] Cox regression – randomised set, [b] estimated as part of a joint frailty model which adjusts for the dependency between the increased risk of death with each subsequent hospitalisation, randomised set; [c] estimated as the cumulative incidence function censoring non-CV-death as a competing risk, randomised set; [d] MMRM, observed case after discontinuation without imputation for death.

3.3.10.2 Geographical region subgroups

In response to a clarification question, the company provided Figure 11, which shows the results for the primary composite outcome by geographic region. The ERG notes that the results suggest a trend towards less benefit with empagliflozin in the Europe region compared to the overall ITT population and the other regions, although the hazard ratio still favours treatment with empagliflozin over placebo.

Figure 11. Primary composite (CV-death or adjudicated HHF) by region in EMPEROR-Reduced (Reproduced from CQ response, Figure 2)



The ERG also requested the results for the Europe subgroup of EMPEROR-R during the clarification stage because in TA679[6] the ERG preferred the used of the Europe subgroup, although the committee concluded that the data from the full DAPA-HF trial population were acceptable for decision making. The ERG notes that the Europe subgroup in EMPEROR-R was [redacted] white and that the company therefore considers the subgroup not to be representative of the UK population, citing data from the 2011 census that reports the England and Wales population comprises of 86% white, 3.3% black, 7.5% Asian and 3.2% other other.[24] The ITT population of EMPEROR-Reduced was 71%

white, 6.6% black, 18.1% Asian and 4.2% other [7] and the company argues that this is more representative of the UK population. The ERG notes that the other baseline characteristics for the Europe subgroup (Table 55) suggest they have a slightly higher mean age (69.6 years) which is closer to that expected in the UK (approximately 75 years) compared to the ITT population (67.2 years), although the proportion of females (19.4%) is both lower than in the ITT population (23.5%) and the UK population (30-40%). Additionally, the ERG notes that the Europe subgroup appears to comprise of a more severe population compared to the ITT population as the Europe population has a higher baseline median NT-proBNP (Europe 2820pg/ml; ITT 1887pg/ml) and slightly higher proportion of patients with stage III or IV NYHA functional class (Europe 27.6%; ITT 24.9%). The ERG also notes there was a higher proportion of patients with baseline implantable cardioverter-defibrillators (Europe 53.1%; ITT 31.0%) or cardiac resynchronisation therapy (Europe 20.1%; ITT 11.8%) in the Europe subgroup compared to the ITT population. There was also a higher proportion of patients with atrial fibrillation in the Europe subgroup (48.6%) compared to the ITT population (35.6%). However, the mean percentage LVEF, HF hospitalization within 12 months and baseline use of heart failure medication were similar to the ITT population (Table 55).

The results of the Europe subgroup for the primary composite outcome, total HHF and CV mortality all consistently show less benefit with empagliflozin compared to the overall EMPEROR-R population (Table 26). As discussed earlier, the ERG considers it important to highlight that the study was not powered to detect statistically significant differences in subgroups and thus caution should be taken in drawing conclusions from the subgroup results. Nevertheless, the ERG considers it important to highlight that there may be a difference in efficacy with empagliflozin related to geographic region.

Table 26. Results of the primary composite outcome, CV-death and total HHF for the Europe subgroup and overall population from EMPEROR-R (Adapted from CQ response, Table 14)

Endpoint	Europe subgroup [25] N with event / N analysed		Europe subgroup [25] Hazard ratio (95% CI)	Overall population [25] N with event / N analysed		Overall ITT population [25] Hazard ratio (95% CI)
	Placebo	Empagliflozin		Placebo	Empagliflozin	
Time to the first event of adjudicated CV-death or HHF	149/677	140/676	0.94 (0.74 to 1.18)	462/1687	361/1863	0.75 (0.65 to 0.86), p<0.001

Total HHF (first and recurrent)	152/677 [a]	144/676 [a]	0.96 (0.70 to 1.33) [a]	553/1867	388/1863	0.70 (0.58 to 0.85), p<0.001
CV-death	72/677	71/676	0.98 (0.71 to 1.36)	202 (10.8)	187 (10.0)	0.92 (0.75 to 1.12)

[a] Total HF hospitalisation event rates were derived from an unadjusted negative binomial model

Abbreviations: CI, confidence interval; CV, cardiovascular; HHF, hospitalisation for heart failure; ITT, intention to treat

3.3.11 Safety

Safety data are reported for the duration of patient participation within the EMPEROR-R trial. Median exposure to study medication was approximately 14 months in both treatment groups, with 61% of patients treated for at least 1 year. A similar overall proportion of patients in the empagliflozin and placebo groups reported at least one adverse event (AE), most of which were of mild or moderate intensity (Table 27).

In EMPEROR-R treatment discontinuation was classified by either a non-fatal adverse event, request by patient or other reasons. A total of [redacted] out of the [redacted] patients assigned to empagliflozin discontinued treatment [redacted] whereas [redacted] patients out of the [redacted] patients assigned to placebo discontinued treatment [redacted] During clarification, the company confirmed that a further [redacted] patients on empagliflozin and [redacted] patients on placebo discontinued due to a fatal adverse event.

The overall frequency of serious AEs was lower in the empagliflozin group [redacted] than in the placebo group [redacted] with the most frequent serious AE being cardiac disorders, reported in [redacted] of the empagliflozin group and [redacted] of the placebo group. There were also fewer participants experiencing renal and urinary disorders in the empagliflozin arm [redacted] compared to those allocated to receive placebo [redacted]

Table 27. EMPEROR-R (overall population) - Summary of AEs (adapted from the CS, Table 30)

Category of AEs	Empagliflozin 10 mg, N (%)	Placebo, N (%)
Number of patients, N (%)	1863 (100.0)	1863 (100.0)
Patients with any AEs	[redacted]	[redacted]
Mild	[redacted]	[redacted]
Moderate	[redacted]	[redacted]
Severe	[redacted]	[redacted]
Investigator-defined drug-related AEs	[redacted]	[redacted]
AEs leading to discontinuation of study medication	[redacted]	[redacted]

Serious AEs	[redacted]	[redacted]
Serious AEs		
Resulting in death	[redacted]	[redacted]
Life threatening	[redacted]	[redacted]
Persistent or significant disability/incapacity	[redacted]	[redacted]
Requires or prolongs hospitalisation	[redacted]	[redacted]
Abbreviations: AE, adverse event; CS, company submission.		

Adverse events of special interest (AESIs) were pre-specified in the protocol as hepatic injury, decreased renal function, ketoacidosis, and AEs leading to lower limb amputation and specific AEs were defined as urinary and genital tract infections, volume depletion and hypotension, confirmed hypoglycaemic events, bone fractures and urinary tract malignancies. The most frequently reported specific adverse events and adverse events of specific interest were acute renal failure, volume depletion, and hypotension, with [redacted] and [redacted] of patients in the empagliflozin arm and [redacted] and [redacted] of patients in the placebo arm, respectively (Table 28). A greater number of participants allocated to empagliflozin experienced uncomplicated genital infection [redacted] compared to those in the placebo group [redacted]. However, no notable difference between groups was observed for complicated genital infections or those leading to treatment discontinuation. There was no notable difference between groups for investigator-defined drug-related AEs ([redacted] in both groups). AE's that were included in the economic model were: urinary tract infection, genital mycotic infection, acute renal failure, hepatic injury, volume depletion, hypotension, hypoglycaemic event and bone fracture (See Section 4 for further details).

Table 28. EMPEROR-R (overall population) - Specific AEs and AE of specific interest (adapted from the CS, Table 32)

AE	Empagliflozin 10 mg, N (%)	Placebo, N (%)
Number of patients	1863 (100.0)	1863 (100.0)
Acute renal failure	[redacted]	[redacted]
Hepatic injury	[redacted]	[redacted]
Ketoacidosis	[redacted]	[redacted]
AEs leading to LLA up to trial completion (investigator-defined)	[redacted]	[redacted]
Urinary tract infection	[redacted]	[redacted]
Genital infection	[redacted]	[redacted]
Volume depletion	[redacted]	[redacted]
Hypotension	[redacted]	[redacted]
Symptomatic hypotension (investigator-defined)	[redacted]	[redacted]

Confirmed hypoglycaemic events	[redacted]	[redacted]
In patients with T2DM	[redacted]	[redacted]
In patients with pre-diabetes	[redacted]	[redacted]
In patients without diabetes or pre-diabetes	[redacted]	[redacted]
Bone fracture	[redacted]	[redacted]
Urinary tract malignancy up to trial completion	[redacted]	[redacted]

Abbreviations: CS, company submission; AE, adverse event; LLA, lower limb amputation; T2DM, type II diabetes mellitus

3.4 Critique of the indirect treatment comparison and meta-analyses

3.4.1 Critique of the trials included in the indirect treatment comparison and meta-analyses

As discussed in Section 3.1, DAPA-HF[16] was identified in the company's SLR and used in a Bucher ITC along with EMPEROR-R[7] to estimate the relative efficacy of empagliflozin versus dapagliflozin in adult patients with symptomatic, but stable HFrEF. In addition, the ERG notes that the company conducted a pooled meta-analysis of SGLT2i vs SoC using EMPEROR-R and DAPA-HF, and the company consider the results to provide supportive data for the results of the ITC.

The company provided a summary table in the CS to give an overview of the EMPEROR-R and DAPA-HF trials (Table 56). DAPA-HF compared dapagliflozin (10 mg or 5 mg once a day) plus SoC with placebo plus SoC. Both EMPEROR-R and DAPA-HF were phase III, multinational, double-blind, randomised controlled trials. The median duration of follow-up in EMPEROR-R (16 months) was slightly shorter than that in DAPA-HF (18.2 months).

EMPEROR-R and DAPA-HF included patients with chronic HFrEF and LVEF≤ 40%. However, in EMPEROR-R, the NT-pro-BNP inclusion criteria differed (varying from ≥ 600 pg/ml to ≥ 5,000 pg/ml) depending on:

- the reduction in ejection fraction (EF);
- prior HHF; and
- whether the patient had atrial fibrillation (AF).

In DAPA-HF, just AF and prior HHF were used to determine NT-pro-BNP inclusion criteria and the minimum thresholds for inclusion only varied from ≥ 600 pg/ml to ≥ 900 pg/ml (CS, Table 25). The resulting baseline characteristics in EMPEROR-R (Table 29) suggest it contains patients with more

severe disease compared to in DAPA-HF (lower LVEF [~27% versus ~31%, respectively] and higher NT-pro-BNP [~1900pg/ml versus ~1400 pg/ml, respectively]). However, there was a higher proportion of patients in NYHA class II in EMPEROR-R compared to in DAPA-HF (~75% vs ~68%), and fewer NYHA class III patients in EMPEROR-R than in DAPA-HF (~24% vs ~32%). The ERG's clinical experts reported that NYHA is subjective and therefore may not fully represent the disease severity of patients. The ERG considers that based on the LVEF and NT-pro-BNP baseline data, EMPEROR-R patients were likely to be sicker than DAPA-HF and at increased risk of HHF and mortality.

It is not possible to compare the prior HHF data; for EMPEROR-R HHF data were restricted to events in the preceding 12 months, whereas DAPA-HF did not report equivalent data. Baseline eGFR was also noted by the company to be lower in EMPEROR-R (~61 ml/min/1.73m2) compared to DAPA-HF (~66 ml/min/1.73m2). Mean age and sex were similar between the two studies (Table 29).

Table 29. Comparison of baseline characteristics of subjects enrolled in EMPEROR-Reduced and DAPA-HF trials (Reproduced from CS, Table 26)

	EMPEROR-Reduced		DAPA-HF	
Treatment (N)	**Empagliflozin (N = 1,863)**	**Placebo (N = 1,867)**	**Dapagliflozin (N = 2,373)**	**Placebo (N = 2,371)**
Age, mean (SD)	67.2 (10.8)	66.5 (11.2)	66.2 (11.0)	66.5 (10.8)
Female sex, n (%)	437 (23.5)	456 (24.4)	564 (23.8)	545 (23.0)
North America, n (%)	212 (11.4)	213 (11.4)	335 (14.1)	342 (14.4)
South/Latin America, n (%)	641 (34.4)	645 (34.5)	401 (16.9)	416 (17.5)
Europe, n (%)	676 (36.3)	677 (36.3)	1,094 (46.1)	1,060 (44.7)
Asia Pacific, n (%)	248 (13.3)	245 (13.1)	543 (22.9)	553 (23.3)
NYHA I, n (%)	0	0	0	0
NYHA II, n (%)	1399 (75.1)	1401 (75.0)	1,606 (67.7)	1,597 (67.4)
NYHA III, n (%)	455 (24.4)	455 (24.4)	747 (31.5)	751 (31.7)
NYHA IV, n (%)	9 (0.5)	11 (0.6)	20 (0.8)	23 (1.0)
LVEF – %, mean (SD)	27.7 (6.0)	27.2 (6.1)	31.2 (6.7)	30.9 (6.9)
NT-pro-BNP – pg/ml, median (IQR)	1,887 (1077, 3429)	1,926 (1153, 3525)	1,428 (857, 2,655)	1,446 (857, 2,641)
Ischaemic HF, n (%)	983 (52.8)	946 (50.7)	1316 (55.5)	1358 (57.3)
HHF, n (%)	577¶ (31.0)	574¶ (30.7)	1,124 (47.4)	1,127 (47.5)
Atrial fibrillation, n (%)	664 (35.6)	705 (37.8)	916 (38.6)	902 (38.0)
Diabetes mellitus, n (%)	927 (49.8)	929 (49.8)	993 (41.8)	990 (41.8)
eGFR – ml/min/1.73m^2 Mean (SD)	61.8 (21.7)	62.2 (21.5)	66.0 (19.6)	65.5 (19.3)

Abbreviations: AF, atrial fibrillation; eGFR, estimated glomerular filtration rate; HF, heart failure; HHF, hospitalisation for heart failure; LVEF, left ventricular ejection fraction; NT-pro-BNP, N-terminal pro hormone B-type natriuretic peptide; NYHA, New York Heart Association. Rows highlighted in pink describe baseline characteristic with significant variation between the two trials.

¶In EMPEROR-Reduced, the number of HHF refers to the previous 12 months, while there was no time limit on prior HHF in DAPA-HF.

Baseline use of ACEi (or ARB) and sacubitril valsartan (ARNi) also differed between EMPEROR-R and DAPA-HF (Table 30). The ERG's clinical experts reported that the baseline use of SoC drugs in DAPA-HF was likely to be more reflective of clinical practice.

Table 30. Standard of care received at baseline in EMPEROR-Reduced and DAPA-HF trials (Reproduced from CS, Table 24)

SoC at baseline	EMPEROR-Reduced		DAPA-HF	
	Empagliflozin	Placebo	Dapagliflozin	Placebo
Diuretic	94.2%	95.9%	93.4%	93.5%
ACE inhibitor or ARB	70.5%	68.9%	84.2%	82.4%
Sacubitril valsartan	18.3%	20.7%	10.5%	10.9%
Beta-blocker	94.7%	94.7%	96.0%	96.2%
MRA	70.1%	72.6%	71.5%	70.6%
Digitalis	15.2%	16.7%	18.8%	18.6%
Implantable cardioverter-defibrillator	31.0%	31.8%	26.2%	26.1%
Cardiac resynchronisation therapy	11.8%	11.9%	8.0%	6.9%

Abbreviations: ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blocker; MRA, mineralocorticoid receptor antagonist.

3.4.2 Results of the Bucher ITCs

The results of the Bucher ITCs showed [redacted] between empagliflozin and dapagliflozin (Table 31). However, there was a trend suggesting [redacted] CV deaths and all-cause mortality with empagliflozin compared to dapagliflozin (HR [empagliflozin vs dapagliflozin] [redacted] respectively). In addition, there appears to be a [redacted] mean change in KCCQ-TSS score from baseline at 8 months with dapagliflozin (mean difference [MD] [redacted]) compared with empagliflozin (MD [redacted]), albeit potentially [redacted] The ERG notes that neither study was

powered to detect differences in these specific outcomes as they were both powered for their own trial specific primary composite outcomes.

The renal composite outcome was defined differently in the two studies (CQ response, Table 6), with the definition in DAPA-HF being time to the first occurrence of:

- ≥ 50% sustained decline in eGFR;
- reaching end-stage renal disease (ESRD; ESRD was defined as a sustained [≥28 days] eGFR of <15 ml per minute per 1.73 m2, sustained dialysis, or renal transplantation); or
- renal death.

The Bucher ITC of worsening renal function using the DAPA-HF definition suggests that empagliflozin might be [redacted] in reducing the hazard of worsening renal function compared to dapagliflozin, although the [redacted] (HR [redacted]; 95% CI: [redacted]).

Table 31. Summary of Bucher ITC results for empagliflozin plus SoC versus dapagliflozin plus SoC (EMPEROR-Reduced vs DAPA-HF, ITT population) (Reproduced from CS, Table 23)

Endpoint: relative effect measure	EMPEROR-REDUCED: empagliflozin versus placebo[a]	DAPA-HF: dapagliflozin versus placebo[a]	Bucher ITC: empagliflozin versus dapagliflozin[a]
Time to first event of adjudicated CV death or adjudicated HHF: HR (95% CI)	0.75 (0.65 to 0.86)	0.75 (0.65 to 0.85)	[redacted]
Time to first event of adjudicated CV death or adjudicated HHF (EMPEROR-Reduced) versus Time to first worsening of heart failure (hospitalisation or an urgent visit resulting in intravenous therapy for heart failure) or CV death (DAPA-HF): HR (95% CI)	0.75 (0.65 to 0.86)	0.74 (0.65 to 0.85)	[redacted]
Time to first adjudicated HHF: HR (95% CI)	0.69 (0.59 to 0.81)	0.70 (0.59 to 0.83)	[redacted]
Time to adjudicated CV death: HR (95% CI)	0.92 (0.75 to 1.12)	0.82 (0.69 to 0.98)	[redacted]
Time to all-cause mortality: HR (95% CI)	0.92 (0.77 to 1.1)	0.83 (0.71 to 0.97)	[redacted]
Occurrence of adjudicated HHF (first and recurrent) – analysed using a joint frailty model: HR (95% CI)	0.70 (0.58 to 0.85)	0.71 (0.61 to 0.82)	[redacted]

Occurrence of adjudicated HHF (first and recurrent) – analysed using a Lin-Wei-Yang-Ying model: RR (95% CI)	0.76 (0.65 to 0.89)	0.75 (0.65 to 0.88)	[redacted]
Worsening renal function (as defined in DAPA-HF): HR (95% CI)	0.52 (0.29 to 0.92)	0.71 (0.44 to 1.16)	[redacted]
Change in KCCQ total symptom score at 8 months/7.4 months: MD (SE/95% CI)	1.6 (0.7)	2.8 (0.5)	[redacted]

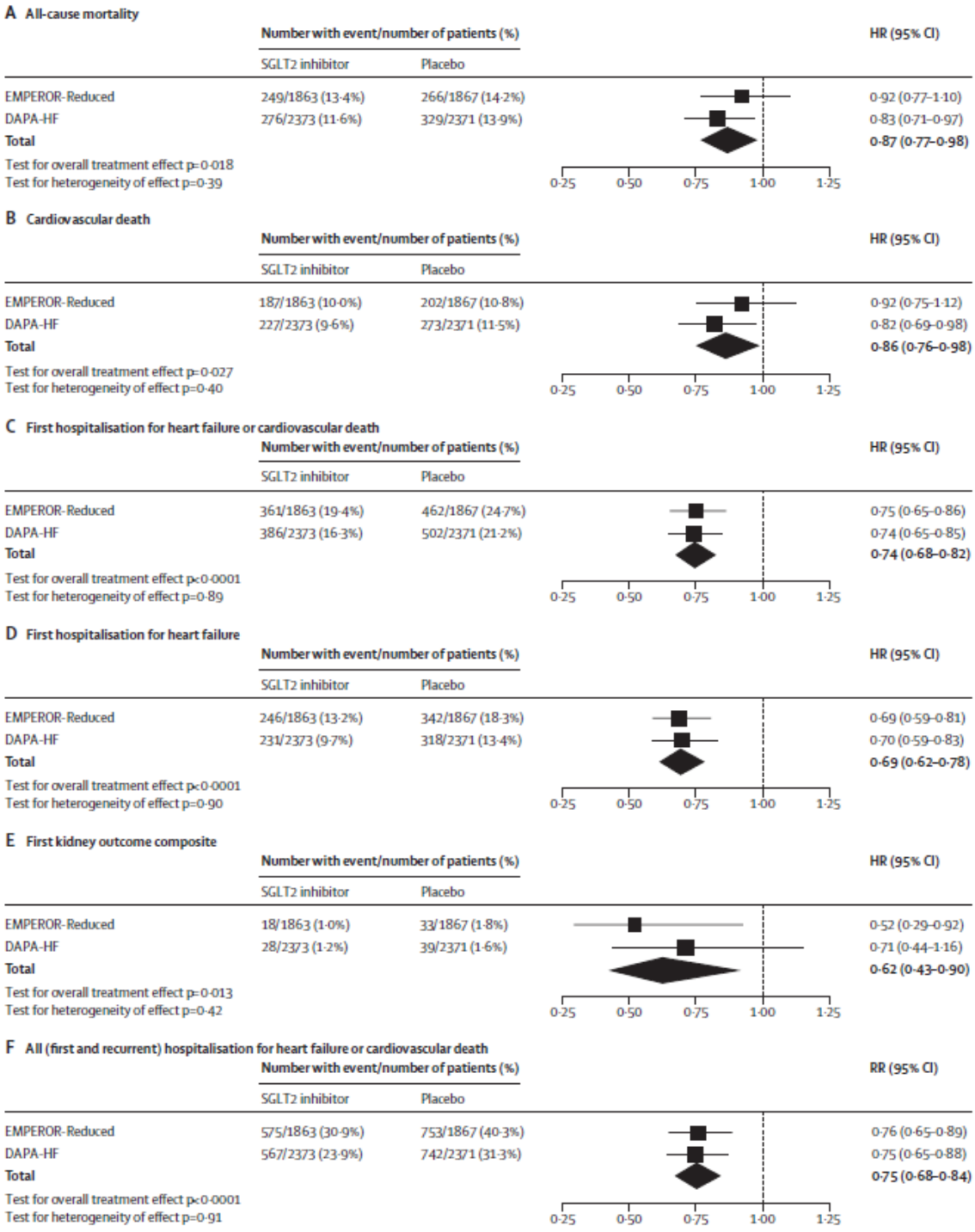
Abbreviations: CV, cardiovascular; HHF, hospitalization for heart failure; HR hazard ration; CI, confidence interval; RR, risk ratio; KCCQ, Kansas City Cardiomyopathy Questionnaire.

3.4.3 Results of the pooled meta-analyses

The company included the results of a published pooled meta-analysis reported by Zannad *et al*. 2020[26] in the CS, where dapagliflozin and empagliflozin data were pooled to generate estimates of effect for SGLT2i versus placebo. The company reported that they used a fixed effects model for the meta-analysis and the R statistical software was used to conduct the analyses. However, the ERG is concerned that the company's assumption of equal effectiveness for SGLT2is is based on only a single trial for empagliflozin and for dapagliflozin and therefore considers the company's Bucher ITC a more appropriate method of assessing the comparative efficacy of empagliflozin and dapagliflozin for this appraisal.

The results of the Zannad *et al*. pooled meta-analysis are presented in Figure 12 and show a statistically significant benefit with the SGLT2is compared to placebo for all the outcomes reported. However, as discussed in Section 3.4.2, there was a notable difference in results between EMPEROR-R and DAPA-HF for the outcomes of all-cause mortality, CV mortality and renal composite outcome.

Figure 12. Meta-analysis of EMPEROR-Reduced and DAPA-HF trials (ITT population) (Reproduced from CS Figure 17)



3.5 Critique of PULSE study

3.5.1 Summary of PULSE study methodology and patient population

The PULSE study (Incidence Prevalence and resoUrce utiLiSation of hEart failure in England) was a non-interventional cohort study based on existing data from the UK Clinical Practice Research Datalink (CPRD) linked to the Hospital Episodes Statistics (HES) inpatient and the Office for National Statistics (ONS) mortality data.[5] The PULSE study was used by the company to check the external validity of the economic model and to inform the modelling of mortality.

PULSE included adults (aged 18+), followed in primary care practices in England, who contributed to the [redacted] and had been diagnosed with chronic HF between 1 January 2015 and 31 December 2019 in either primary or secondary care. In total, PULSE comprised [redacted] patients, although only [redacted] had an identifiable ejection fraction subtype at or prior to the study index date of HFrEF. The data from the reduced ejection fraction (rEF) subgroup were the only data from PULSE used for the economic model and therefore from here on the use of the term PULSE will refer to only the rEF subgroup of the PULSE study. Patients in PULSE have received treatments currently used in UK clinical practice for HFrEF and therefore the data from PULSE is likely to be the equivalent of SoC. However, the ERG notes that empagliflozin is recommended for treatment of patients with symptomatic HFrEF and clinical experts reported that upto 30% of HFrEF patients may be asymptomatic (NYHA Class 1). PULSE [redacted] but given the likely small proportion of patients who are asymptomatic and the absence of other data, the ERG considers the use of the HFrEF subgroup data from PULSE to be reasonable.

The baseline characteristics of the PULSE study are summarised alongside those for the placebo arm of EMPEROR-R in Table 32. The ERG notes that patients in the placebo arm of EMPEROR-R were approximately [redacted] years younger and therefore patients from PULSE are [redacted] with the ERG's clinical experts estimate of the age of patients likely to be eligible for empagliflozin. The proportion of patients with HHF in the previous 12 months (≤12 months) was [redacted] in EMPEROR-R (30.7%) compared to in PULSE ([redacted]%) [redacted] the targeted inclusion criteria of EMPEROR-R to identify patients at increased risk of adverse HF outcomes. However, the company also report that the HHF events in EMPEROR-R were independently adjudicated, whereas in PULSE this was not the case, and so it is possible that there may be some

errors in coding of hospitalisation events for some patients with HHF incorrectly recorded for non-HHF events and vice versa. The extent of these potential coding errors in PULSE is unknown and therefore it is not possible to predict the direction of any resulting bias.

The ERG also notes that there was a [redacted] median NT-ProBNP (1,926 vs [redacted]) and [redacted] mean LVEF (27.2 vs [redacted]) in EMPEROR-R compared to in PULSE, although the data from PULSE are based on [redacted]% of patients due to missing data. Nevertheless, [redacted] compared to PULSE. Data on NHYA classification in PULSE were [redacted] of patients, but the ERG notes PULSE [redacted] patients in EMPEROR-R. Data on baseline use of MRA and sacubitril valsartan were also reported to be [redacted] The available data suggest a [redacted] proportion of EMPEROR-R received a BB, MRA or sacubitril valsartan than the population in PULSE, which is also likely to reflect [redacted]

The ERG notes that the main determinants of risk in the economic model were the KCCQ-CSS quartiles from EMPEROR-R and, unfortunately as KCCQ scores were not available in the CPRD, the KCCQ-CSS for the PULSE population is unknown.

Table 32. Baseline characteristics of PULSE HFrEF cohort (incident and prevalent) and EMPEROR-reduced (Adapted from CS Appendix O, Table 3 and CQ response, Table 17)

Patient characteristics	PULSE (n=68,780)	EMPEROR PBO (n=1,867)
Age at index (years)	[redacted]	66.5
Sex, females (%)	[redacted]	24.40%
Time since HF Diagnosis (years)	[redacted]	6.3
Hospitalisation for HF in prior 12 months	[redacted]	30.7%
NYHA classification		
Stage 1	[redacted]	0%
Stage 2	[redacted]	75.0%
Stage 3	[redacted]	24.4%
Stage 4	[redacted]	0.6%
Missing	[redacted]	0%
BMI (kg/m^2), mean	[redacted]	27.8
LVEF, mean	[redacted]	27.2

NT-ProBNP, median	[redacted]	1,926
Systolic Blood Pressure, mean	[redacted]	121.4
Heart Rate (bpm), mean	[redacted]	71.5
eGFR (ml/min/1.73m^2)	[redacted]	62.2
Type 2 diabetes (%)	[redacted]	49.80%
Past history of ischemic heart disease (%)	[redacted]	50.7%[c]
ACEi/ARB	[redacted]	68.9%
Beta-Blocker	[redacted]	94.7%
MRA	[redacted]	72.6%
Sacubitril Valsartan	[redacted]	20.7%

[redacted]

Abbreviations: BMI, body mass index; bpm, beats per minute; HF, heart failure; PBO, Placebo; LVEF, Left Ventricular Ejection Fraction; NT-ProBNP, N-terminal pro-brain natriuretic peptide; eGFR, estimated glomerular filtration rate; NYHA, New York heart association; MRA, Mineralocorticoid receptor antagonists.

3.5.2 Results from PULSE

In response to clarification, the company provided results from PULSE and EMPEROR-R as rates per patient year to enable a comparison between the studies, along with supportive figures with the results separately for each outcome from each study (CQ response, Figures 3 to 8). As noted in Section 3.5.2, HHF events in EMPEROR-R were adjudicated, whereas in PULSE they weren't and thus PULSE could be at risk of reporting bias. Likewise, CV-death was an adjudicated endpoint in EMPEROR-R, whereas in PULSE, the recording of CV death was dependent on the accuracy of the ONS records. However, the impact of this potential bias in PULSE is unknown and so the results should be interrupted with caution.

The cumulative number of HHF events (first and recurrent) was [redacted] for PULSE than EMPEROR-R with an age and sex adjusted HHF rate of [redacted] patient years in PULSE compared to [redacted] patient years in the placebo arm of EMPEORER-R. The age adjusted CV-death rate for HFrEF [redacted] in PULSE ([redacted] patient years) compared to in the EMPEROR-R placebo arm ([redacted] patient years). [redacted], the all-cause mortality rate in PULSE was [redacted] ([redacted] patient year [age and sex adjusted]) compared to in EMPEROR-R ([redacted] patient years).

The company concluded that, "*although there might be differences in the baseline characteristics between the trial population and the UK clinical practice, the all cause-death outcomes observed in PULSE and EMPEROR-Reduced were broadly comparable, indicating that the difference in HHF rates might be due to how they were recorded*." However, the ERG considers the difference in HHF may be a result of the increased disease severity of patients in EMPEROR-R. For further critique of the PULSE study and the validation of the economic model please see Section 4.

3.6 Conclusions of the clinical effectiveness section

The ERG considers the company's SLR to be of reasonable quality and likely to have retrieved all studies relevant to empagliflozin, despite limiting inclusion to English-language publications. The ERG also considers EMPEROR-R, the key study informing the clinical effectiveness of empagliflozin, to be a well-designed and well-conducted RCT, with an overall low risk of bias and high internal validity. However, the ERG has concerns about the external validity of EMPEROR-R and its applicability to patients in clinical practice in England with HFrEF who are likely to be eligible for treatment with empagliflozin. Firstly, the ERG notes that the inclusion criteria of EMPEROR-R were specifically designed to recruit patients with an increased risk of an outcome event and who had a markedly reduced ejection fraction and increased levels of natriuretic peptides. The ERG considers that this is reflected in the baseline characteristics of the trial, with patients having a low ejection fraction and high NT pro-BNP. The ERG considers that as a result of the increased severity of patients in EMPEROR-R, the patients in the trial are likely to have an increased risk of poorer health outcomes than expected in clinical practice and therefore the generalisability of the trial results is unknown.

As is typical of randomised controlled trials, the ERG notes that the population of EMPEROR-R comprises of a younger population and higher proportion of males than in clinical practice. Additionally, the ERG's clinical experts reported that the proportions of some of the SoC drugs used in EMPEROR-R were not completely reflective of current clinical practice in England. In particular, the clinical experts reported that they would expect a slightly lower proportion of patients to be taking sacubitril valsartan than the ~20% in EMPEROR-R and around 10% more patients to be taking an ACEi or ARB (~70% in EMPEROR-R). However, the ERG notes that SoC drugs use at baseline were reasonably well balanced between the trial arms in EMPEROR-R.

For the primary composite outcome of CV death or HHF in EMPEROR-R, the company conducted a Cox regression adjusted for age, baseline estimated glomerular filtration rate (eGFR; based on the serum creatinine Chronic Kidney Disease Epidemiology Collaboration equation [(CKD-EPI)cr][18]),

region, gender, treatment, baseline diabetes status and LVEF which showed a statistically significant reduction in the risk of CV death or HHF with empagliflozin compared with placebo (HR 0.75; 95% CI 0.65 to 0.86, p<0.0001). The ERG notes that the result of the primary analysis was consistent with that of the sensitivity analyses. The ERG also notes from the plots for the individual outcomes of CV death and first HHF that the benefit for empagliflozin appears to be driven by the HHF events.

The results for the analyses of renal outcomes showed the adjusted mean eGFR change from baseline was 3.3 mL/min/1.73m^2 (95% CI: 1.8 to 4.8) for empagliflozin versus placebo. The ERG notes that this is lower than the 5 or greater mL/min/1.73m^2 minimal clinically meaningful difference observed in other studies (e.g. Mayne *et al.*)[21] but the ERG's clinical experts consider that it is likely that there will be a long-term renal benefit with empagliflozin. In addition, the ERG notes that EMPEROR-R was not powered to detect a difference in renal outcomes and thus considers the results should be interpreted with caution.

The ERG notes that the KCCQ-CSSs were used to inform transition probabilities in the economic model. The KCCQ-CSS adjusted mean difference between empagliflozin and placebo was [redacted] at weeks 12, 32 and 52 and [redacted] (Table 23). [redacted] the adjusted mean difference between empagliflozin and placebo in KCCQ-CSS score was [redacted] at each timepoint and that published literature suggests a minimum of a 5 point difference in the KCCQ overall score is clinically significant.[23]

Direct head-to head data for empagliflozin versus dapagliflozin in HFrEF are not available, and to generate estimates of effect required that the company carry out an ITC. In the CS, the company presented a Bucher indirect comparison using EMPEROR-R and DAPA-HF with placebo (plus SoC) as the common comparator. The ERG considers that based on the LVEF and NT-pro-BNP baseline data, EMPEROR-R patients were likely to be sicker than those in DAPA-HF and at increased risk of HHF and mortality. However, in the absence of other more suitable data the ERG considers it reasonable to conduct a Bucher ITC of the two trials.

The results of the Bucher ITC showed [redacted] between empagliflozin and dapagliflozin. However, there was a trend suggesting [redacted] CV deaths and all-cause mortality with empagliflozin compared to dapagliflozin (HR [empagliflozin vs dapagliflozin] [redacted], respectively) and a [redacted] mean change in KCCQ-

TSS score from baseline at 8 months with dapagliflozin (mean difference [MD] [redacted]) compared with empagliflozin (MD [redacted]).

The ERG notes that the results of the ITC are not used in the economic model, and the company instead assumes equal clinical-effectiveness for empagliflozin and dapagliflozin. The ERG is concerned that the company is making a strong assumption of equivalence for empagliflozin and dapagliflozin based on a single trial for each drug, with [redacted] results from the ITC. The ERG thus considers the results of the pooled meta-analysis conducted by the company, where it is assuming a class effect for SGLT2is, should be interpreted with caution.

The ERG considers the subgroup analyses of patients by age and geographic region in EMPEROR-R suggest there may be differences in the treatment effect with empagliflozin. The results for the primary composite outcome of time to first event of adjudicated CV death or adjudicated HHF suggest patients aged [redacted] years have a [redacted] benefit than those aged [redacted] years and the results for the overall population may be being driven by the [redacted] subgroup. For the geographic region subgroup, patients from Europe had the least benefit with empagliflozin versus placebo compared to the other geographic regions. The ERG appreciates that the study was not powered to detect statistically significant differences in subgroups but considers the Europe subgroup is relatively large (n=1,353)

The PULSE study was used by the company to check the external validity of the economic model and to inform the modelling of mortality. The ERG considers the population in PULSE to be more representative of the HFrEF population in England likely to receive empagliflozin compared to the population in EMPEROR-R. However, data from PULSE are also subject to limitations as they were not adjudicated and incorporate data from asymptomatic HFrEF patients.

Data from PULSE show a [redacted] rate of HHF compared to with SoC in EMPEROR-R (age and sex adjusted HHF rate of [redacted] patient years in PULSE compared to [redacted] patient years in the placebo arm of EMPEOREER-R). However, unfortunately KCCQ-CSS scores were not collected in PULSE and the company's model structure relies on KCCQ-CSS scores to estimate HHF. The company therefore reported that they could not incorporate the data from PULSE in the economic model (Section 4.1.6.3).

4 Cost effectiveness

The company carried out a systematic literature review (SLR) using separate search strategies to identify existing cost-effectiveness evidence, health-related quality of life (HRQoL) data, and cost and healthcare resource use for patients with heart failure with reduced ejection fraction (HFrEF). The searches were initially run in April 2020 and updated in October 2020. The electronic database searches had no date limit and conference literature was searched from January 2018.

A summary of the ERG's assessment of the company's SLR is presented in Table 33. Due to time constraints, the ERG was unable to replicate the company's searches and appraisal of identified abstracts.

Table 33. Summary of SLR

Systematic review step	Section of CS in which methods are reported			ERG assessment of robustness of methods
	Cost effectiveness evidence	HRQoL evidence	Resource use and costs evidence	
Search Strategy	Appendix G. Tables 1-7.	Appendix H. Tables 1-7.	Appendix I. Tables 1-7.	Appropriate. The company searched MEDLINE, Embase, MEDLINE-in-progress, HTAD and NHS EED, and Econlit, plus hand searched grey literature covering three years of conference abstracts from AHA, ESCC, ISPOR and ACC, plus NICE, CADTH, SMC, AWMSG and IQWIG.
Inclusion/Exclusion criteria	Appendix G. Table 8.	Appendix H. Table 8.	Appendix I. Table 8.	Appropriate. PICOS based approach is clear and comprehensive.
Screening	Appendix G. Figure 1.	Appendix H. Figure 1.	Appendix I. Figure 1.	Appropriate. PRISMA flow diagram provided.

Data Extraction	Appendix G. Page 18.	Appendix H. Page 11.	Appendix I. Page 11	Appropriate, though example templates were not provided.
Quality assessment of included studies	Appendix G. Tables 10 and 11, Figure 2.	NR	NR	Appropriate. The Drummond and Jefferies checklists were used.

Abbreviations: CS, company submission; ERG, evidence review group; HRQoL, health related quality of life; AHA, American Heart Association; ACC, American College of Cardiology; ESCC, European Society of Cardiology Congress; ISPOR, International Society for Pharmacoeconomics and Outcomes Research; NICE, National Institute of Health and Care Excellence, CADTH, Canadian Agency for Drugs and Technologies in Health; SMC, Scottish Medicines Consortium;

Following full text screening, 44 cost-effectiveness studies were included for analysis, of which nine were conducted in the UK and considered relevant for the development of the economic model. Two studies were NICE health technology assessment (HTA) submissions,[14, 27] three were SMC HTA submissions,[28-30] one was part of the All Wales Medicines Strategy Group (AWMSG) submission,[31] and three consisted of cost-effectiveness studies.[32-34] Sixty-one HRQoL studies and 45 cost studies were included for data extraction.

The company used the HRQoL data from the EMPEROR-R trial to estimate utility values for each Kansas City Cardiomyopathy Questionnaire clinical summary scores (KCCQ-CSS) heath state in the model. One disutility value for adverse events was identified from the literature,[35] as was the study that provided the age-related adjustments for KCCQ-CSS utilities.[1] These studies were not identified in the HRQoL SLR.

Two of the cost studies used as sources in the economic analysis were not identified through the SLR,[36, 37] although Alva *et al.* 2015[38] was listed in the extraction template as a source in the McEwan *et al*. study,[32] identified through the SLR. The sources used to inform cost and resource use data in the model are discussed further in Section 4.1.9.

4.1 Summary and critique of company's submitted economic evaluation by the ERG

4.1.1 NICE reference case checklist

Table 34 summarises the ERG's appraisal of the company's economic evaluation against the requirements set out in the NICE reference case checklist for the base-case analysis, with reference to the NICE final scope outlined in Section 2.

Table 34. NICE reference case checklist

Element of health technology assessment	Reference case	ERG comment on company's submission
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Yes.
Perspective on costs	NHS and PSS	Yes.
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	Yes.
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Yes.
Synthesis of evidence on health effects	Based on systematic review	Yes.
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults.	Yes.
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	Yes.
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	Yes.
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes.
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Yes.
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	Yes.

Abbreviations: ERG, evidence review group; NHS, national health service; PSS, personal social services; QALY, quality adjusted life year

4.1.2 Population

The population in the base case economic model consists of the ITT population from EMPEROR-R, which included adults with symptomatic chronic HFrEF, with or without diabetes.

As discussed in Section 2.3.1, the ERG's clinical experts indicated that the ITT population in EMPEROR-R is reflective of a sicker and younger population than patients seen in the UK clinical practice (typical of clinical trials in heart failure). The experts pointed that the UK HFrEF population is

on average a decade older than in EMPEROR-R (baseline age 67 years); includes a higher proportion of female patients; has much lower NT pro-BNP values at baseline and higher ejection fraction. Another noted crucial difference between the trial and the UK population is the high number of hospitalisations for heart failure (HHF) observed in EMPEROR-R, with experts agreeing that for a more stable population, there are very few (if any) hospitalisations. Both experts agreed that having a sicker trial population served the purpose of capturing more events in order to demonstrate the effect of empagliflozin over the trial period. One expert expressed the view that only higher risk patients would derive a benefit from empagliflozin. The age subgroup data from EMPEROR-R on hospitalisations for patients below 65 years also showed a difference in efficacy compared to the ≥65 years subgroups (discussed further in Section 3.3.10.1).

The clinical experts also noted that the population included in the PULSE study (a retrospective observational study of the burden of HF, including HFrEF, in England (described in detail in Section 3.5) was reflective of patients seen in UK clinical practice.[5] The company has used PULSE results to check the external validity of the economic model outputs and used the mortality in PULSE to base their choice of the Weibull model to estimate long-term mortality in the model.

The clinical experts advising the ERG noted that the long-term survival predictions of the Weibull curves used in the company's base case underestimated the long-term mortality expected to be seen in high-risk HFrEF populations, such as that of the EMPEROR-R trial. The experts added that the curves used by the company were more representative of patients seen in the UK clinical practice and suggested that the long-term predictions of the Gompertz curves would be a better representation of the shorter survival expected in the trial population.

Therefore, the ERG notes that the company's economic model reflects a hybrid of the EMPEROR-R population and the UK population with HFrEF, where the rate of hospitalisations taken from the trial and used in the model reflect a sicker population (and thus prone to more hospitalisations) but the choice of distribution to estimate survival in the model reflects a less sick population, with longer longevity, more closely reflecting UK clinical practice.

During clarification, the ERG requested that the company ran the economic analysis on two subpopulations: 1) a more severe population, reflective of the EMPEROR-R trial (with high HHF and shorter survival); and 2) an older, less severe population, reflective of UK patients with HFrEF (with less HHF and longer survival). For scenario 2, the ERG asked that the company used the PULSE study

to inform and validate the HHF outcomes in the economic model and used the Weibull model to estimate long-term mortality in the analysis.

Scenario 1: Trial population

The company conducted a scenario where the clinical outcomes remained those used in their base case, and where mortality (both overall and CV-related) was estimated with Gompertz curves, reflecting a shorter survival.

The ERG considers that scenario 1 conducted by the company reflects a better representation of the trial population and data than the company's base case. Therefore, to avoid the inconsistency between trial and PULSE input data, the ERG presents results of its scenario analysis in this population to reflect trial results (Section 6).

The ERG notes that although using a Gompertz curve reflects a shorter long-term survival, the company's scenario 1 still overestimates the number of initial deaths in the model when compared to the same period of EMPEROR-R. These issues are discussed in Section 4.1.6.9.

Scenario 2: UK population with HFrEF

The company conducted a scenario analysis (discussed in Section 3.3.10) where the data from the subgroup analysis from EMPEROR-R for patients above 65 years (mean age 74 years) were used. This scenario used the subgroup analysis undertaken by the company to estimate the HHF and mortality risk equations in the model, together with using a Weibull model as requested by the ERG.

As acknowledged by the company, although the results of the subgroup analysis pertain to a population with a similar mean age as that of the patients expected to be treated in the UK, it still reflects a higher risk population than PULSE and the UK population, given the sicker patients included in EMPEROR-R. The company concluded that the model still overestimated CV-related deaths and HHF compared to the PULSE data. The company added that this was due to the EMPEROR-R population having a higher risk of adverse CV outcomes compared to the PULSE cohort.

The company stated that using PULSE HHF data in the model was not feasible because KCCQ-CSS scores were not collected in PULSE and the company's model structure relies on KCCQ-CSS scores to estimate HHF. This issue is further discussed in Section 4.1.4 and Section 4.1.6.3.

In Section 4.1.6.4 and in Section 4.1.6.9.3, the ERG discusses the overestimation of HHF and CV deaths in the model when compared to PULSE in detail. Nonetheless, the ERG notes that modelling a population that is representative of UK's HFrEF patients is of paramount importance to aid the committee's decision making. Importantly, the ERG remains concerned that the company has not provided a scenario analysis which reflects the whole population considered in this appraisal. The analysis conducted by the company for scenario 2 still considerably overestimates HHF in the model when compared to PULSE (and therefore with the UK population).

4.1.3 Interventions and comparators

The intervention included in the economic model was empagliflozin formulated as a 10mg tablet taken once a day, in addition to standard of care (SoC). For simplicity, hereafter, the ERG refers to the intervention as empagliflozin.

Standard of care was modelled as a basket of drugs used in first-line heart failure care. The assumed proportions of each drug used in SoC is given in Table 35. The clinical experts advising the ERG agreed with the drugs included in the company's basket of SoC treatments, however, disagreed with some of the proportions used. This is further discussed in Section 4.2.9.

Table 35 – Composition of Standard of Care

Drug category	Proportion in ITT population
ACEi	45.4%
ARB	24.3%
ARNi	19.5%
MRA	71.3%
Beta blocker	94.7%
Loop or high ceiling Diuretics	84.5%

4.1.3.1 Dapagliflozin

Dapagliflozin was not included as a comparator in the base case model. After a clarification request from NICE, the company provided a cost comparison analysis as it concluded that dapagliflozin and empagliflozin have a similar effectiveness profile.

The company added that the acquisition costs of dapagliflozin and empagliflozin are the same - the list price of empagliflozin and dapagliflozin is £36.59 and both drugs have the same dosing frequency and method of administration. None of the drugs have patient access schemes in place. Therefore, the company concluded that a cost utility analysis (and corresponding ICER) to estimate the cost effectiveness of empagliflozin vs dapagliflozin was not necessary.

As discussed in Section 3.4 and 3.5, the ERG is concerned that the company is making a strong assumption of equivalence for empagliflozin and dapagliflozin based on a single trial for each drug, with [redacted] results from the ITC and therefore, considers the company's Bucher ITC a more appropriate method of assessing the efficacy of empagliflozin vs dapagliflozin for this technology appraisal.

Given the results of the cost comparison analysis are based on the assumption of a similar effectiveness profile between the two drugs, the ERG considers that the analysis conducted by the company is not sufficient to inform the committee on the cost effectiveness of empagliflozin vs dapagliflozin and recommends that the company includes the results of the Bucher ITC in the economic model.

4.1.4 Modelling approach and model structure

4.1.4.1 Model structure

The company developed a cohort state-transition model with five health states (Figure 13). The four KCCQ-CSS health states represent the different levels of disease severity experienced by patients. KCCQ-CSS quartiles 1 to 4 correspond to KCCQ-CSS scores of 0 to <[redacted] respectively, with higher scores corresponding to a better health status. Patients entered the model according to the baseline distribution of KCCQ-CSS quartiles in EMPEROR-R and could transition to a higher or lower KCCQ-CSS quartile; remain in the same state; or die. Patients could have a CV-related death or a non-CV death.

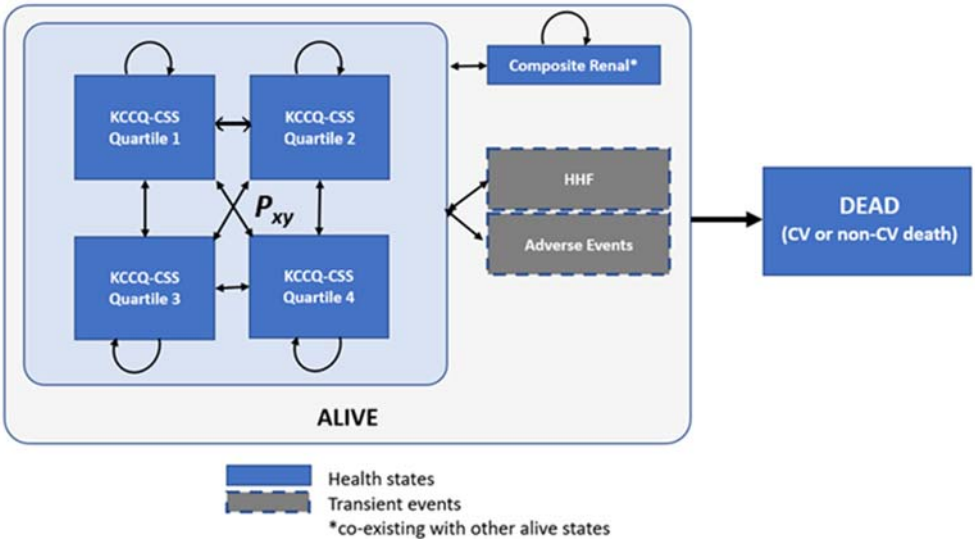
In each of the KCCQ-CSS states, patients had a probability of experiencing an HHF; a treatment-related adverse event (AE); or a composite renal event. The composite renal event included chronic dialysis; renal transplant; and/or sustained reduction in eGFR.

The company's model structure allowed for the estimation of the relationship between disease progression (measured through movements in the KCCQ-CSS states) on outcomes such as HHF;

survival; quality of life; and time on treatment. This was done by introducing different KCCQ-CSS state predictors for each quartile in the HHF; survival; quality of life; and time on treatment equations. Therefore, every time a patient moved KCCQ-CSS states in the model, their probability of HHF or death, and their quality of life also changed.

Patients could discontinue treatment with empagliflozin at any cycle. After discontinuation, patients in the model were assumed to receive SoC until dead.

Figure 13. Company's model structure



Abbreviations: CSS, clinical summary score; CV, cardiovascular; HHF, hospitalisation for heart failure; KCCQ, Kansas City Cardiomyopathy Questionnaire. Q1, Quartile 1: 0 to <55.2; Q2, Quartile 2: 55.2 to <75; Q3, Quartile 3: 75 to <89.6; Q4, Quartile 4: 89.6 to 100.

4.1.4.2 ERG critique

The ERG notes that using time-varying KCCQ-CSS states in the model to estimate the impact of disease progression over time on HHF and survival adds richness to the economic analysis when using the EMPEROR-R trial data. Nonetheless, the ERG also notes that it makes the model less flexible to using alternative data inputs. For example, if HHF and survival were not dependent on patients' transitions through KCCQ-CSS states in the model, the PULSE data could have been directly used to model HHF and survival in a scenario analysis, which would have allowed the model to better reflect patients with HFrEF in the UK.

The company's justification for the choice of thresholds for the KCCQ-CSS states in the model was that these included an adequate number of patients in each category to permit statistically robust

analysis and predicting patient outcomes, when compared to other choices of quartiles. Overall, the ERG is satisfied with the company's choice of KCCQ-CSS states in the model and notes that in the dapagliflozin STA, the company's model used KCCQ scores of 0 to <58, 58 to <77, 77 to <92 and 92 to 100 ([redacted]). In the dapagliflozin FAD, the ERG noted that, "*cut-offs for the quartiles chosen by the company to measure KCCQ-TSS in the model were arbitrary. But it said it expected that using other cut-offs or approaches to grouping would minimally affect the cost-effectiveness results. The committee concluded that the company's model structure was appropriate for decision making*".

4.1.5 Perspective, time horizon and discounting

A lifetime horizon was adopted in the model and time was discretised into monthly cycles with a half-cycle correction applied. The analysis was carried out from an NHS and Personal Social Services (PSS) perspective. Costs and health effects are discounted at an annual rate of 3.5%, in line with the NICE Reference Case.

The ERG agrees with the lifetime horizon used, however, notes that the model adopts a time horizon of 33 years (instead of 60 as reported in the CS). The ERG also notes that patients' baseline age in the company's model was 67 years. Therefore, at 24 years in the model (when patients would be 91 years), 1% of empagliflozin patients are still alive. Given the level of disease severity of the modelled population, the ERG considers this to reflect an overestimation of long-term survival (or reflect a healthier population as discussed in Section 4.2.2). This issue is further discussed in Section 4.2.6.9.

The ERG agrees with the use of the half-cycle correction given the monthly cycle length. However, during clarification the ERG noted that the half-cycle correction was being applied twice in the model, which the company corrected.

4.1.6 Treatment effectiveness

Treatment effectiveness was modelled through patients' change in KCCQ-CSS state, the rate of HHF, the probability of experiencing renal outcomes, and mortality. These are discussed in the next sections in detail.

Analyses of overall survival, CV-related mortality and HHF rates (as well as quality of life and time to treatment discontinuation) involved developing regression equations which included a time-varying predictor for patients' KCCQ-CSS state, with or without a treatment group predictor. The company

noted that as time-varying KCCQ-CSS states were included in the regression models as predictors, it was expected that the treatment effect in the equations would become weaker. Nonetheless, the company decided that for mortality and HHF models to capture any benefits above and beyond changes in health states over time, the treatment effect predictor should be maintained in the models. This decision, however, was not applied in the utility analyses, where the treatment predictor was removed from the regression models. These issues are discussed in the appropriate sections below.

4.1.6.1 Transition between KCCQ-CSS states

The clinical study report (CSR) for EMPEROR-R reported that KCCQ-CSS data were collected at baseline; week 12; week 32; week 52; and/or at the end of treatment and 30 days thereafter. Based on the observed quartile distribution of KCCQ-CSS at trial baseline, the company decided to define the KCCQ-CSS health states in the model as quartiles 1 to 4 corresponding to scores of 0 to <[redacted]; [redacted] to <[redacted]; [redacted] to <[redacted]; and [redacted] to 100, respectively.

The company excluded KCCQ-CSS data collected after week 52 in the EMPEROR-R trial on the basis that the number of available observations beyond that period were considered too few and based on exploratory analyses conducted by the company, which revealed that KCCQ-CSS health states tend to change early after start of treatment and stabilise thereafter.

The company used the last-observation carried-forward (LOCF) imputation method to deal with missing KCCQ-CSS data over the first 52 weeks of the trial. Missing measurements due to early end of follow-up (mostly between week 32 and 52) were observed for 14.2% of patients in the placebo arm and 15.1% in the empagliflozin arm and were not imputed since death status past the end of follow-up was unknown and the distribution of the last known KCCQ-CSS for patients with early end of follow-up was similar to the distribution among observed/imputed data.

The company estimated transition probabilities (TPs) from the KCCQ-trial data between the three periods of trial visits (baseline week 12, 12–32 and 32–52) by treatment arm and assessed the TPs for variation over time. The company concluded that the probability of patients transitioning between KCCQ-CSS states varied across the three time periods in the placebo group, therefore, decided to have three sets of period-specific TPs in the model in each treatment arm.

Each of the six derived matrices was then converted to monthly TPs by finding the p-th root of the observed transition matrix for the trial period (e.g., 12 weeks, 20 weeks). This yielded six sets of

monthly TPs representing progression in the three periods used in the analysis (reported in Table 36 in the CS, page 135). The transition matrices for the last period (week 32+) were used to predict the changes in KCCQ-CSS scores for the rest of model time horizon. To note is that when patients discontinue treatment with empagliflozin in the model, the set of TPs used is that of the SoC patients.

4.1.6.2 ERG critique

The ERG was concerned that the company was excluding two additional time points (end of treatment visit, and follow-up visit) for which KCCQ-CSS data were available. Therefore, during clarification the ERG requested that the company provided the number of observations and mean KCCQ-CSS at week 12, week 32, week 52, end of treatment visit, and follow-up visit without imputed values (Table 36) and with imputed values (Table 37).

The company clarified that visits post-week 52 did not have a fixed scheduled time and that in both the imputed and non-imputed analysis, measurements from day 436 (week 62) were recorded as "post-week 52".

The values reported in Table 36 and in Table 37 suggest that after week 52 there was not a meaningful change in the KCCQ-CSS scores in the placebo nor in the empagliflozin arm of EMPEROR-R. Given the number of observations dropped to about 50%, the ERG agrees with the company's approach of not using post-week 52 KCCQ-CSS data in the base case model.

Nonetheless, the ERG notes the discrepancy between the description of available measures for end of treatment visits and follow-up visits in the CSR and in the company's response. It would have added robustness and reassurance to the company's analysis if the trend in estimates shown in Table 36 was confirmed for patients in their follow-up visit, as opposed to a random point in time after week 52, when patients could still be on treatment. Furthermore, the appropriateness of the LOCF method relies on the validation of the company's statement that the initial increase in KCCQ-CSS is maintained, *and* that the missing observations all occur in the "plateau" part of the observations. The ERG requested the results of the exploratory analyses conducted by the company which revealed that KCCQ-CSS health states tend to change early after start of treatment and stabilise thereafter. The company did not supply these analyses.

Table 36. Mean KCCQ-CSS data (without imputed values) from EMPEROR-R

Time period	Placebo N	Placebo mean (SD)	Empagliflozin N	Empagliflozin mean (SD)
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Week 12	[redacted]	[redacted]	[redacted]	[redacted]
Week 32	[redacted]	[redacted]	[redacted]	[redacted]
Week 52	[redacted]	[redacted]	[redacted]	[redacted]
Post week 52	[redacted]	[redacted]	[redacted]	[redacted]

Table 37. Mean KCCQ-CSS data (with imputed values) from EMPEROR-R

Time period	Placebo N	Placebo mean (SD)	Empagliflozin N	Empagliflozin mean (SD)
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Week 12	[redacted]	[redacted]	[redacted]	[redacted]
Week 32	[redacted]	[redacted]	[redacted]	[redacted]
Week 52	[redacted]	[redacted]	[redacted]	[redacted]
Post week 52	[redacted]	[redacted]	[redacted]	[redacted]
* the higher number of observations at week 12 are due to records from patients with missing scores at baseline. These patients contribute data on transitions from week 12 onwards and were kept in the analyses.				

The ERG notes that the company's methodology to estimate monthly TPs by finding the p-th root of the observed transition matrix for the trial period (e.g., 12 weeks, 32 weeks, etc.) would only be appropriate if the observed matrices from the trial had non-singular roots. As the company did not share the original matrices, the ERG has no way of confirming the non-singularity of the matrices and the appropriateness of the method used.

Overall, the ERG is concerned that there is a disconnection between the reported KCCQ-CSS results from the trial and the input KCCQ-CSS data used the model. While the trial only reported mean KCCQ-CSS values (and changes) over time, the model only made use of patients' distribution and movements across the four KCCQ-CSS quartiles defined by the company, therefore making the validation of the model results against the trial extremely difficult. During clarification, the ERG asked that the company to provide the overall change in mean KCCQ-CSS estimated in the model (i.e., considering the changes from the baseline mean KCCQ-CSS in each quartile) for month 3; month 8; and month 12 and compared it to the mean changes for empagliflozin and SoC observed in EMPEROR-R for the same time points.

The company replied that such comparison would not be possible as the model tracks the change in proportion of patients in each KCCQ-CSS state instead of change in the mean outcome. The company supplied Table 35 (reported in the clarification question document, question B12), which provided the observed proportion of patients in each KCCQ-CSS quartile at weeks 12, 32, and 52 in EMPEROR-

R versus the model proportions at the same times. The company concluded that the observed and predicted KCCQ-CSS proportions were closely matched in the trial and in the model. The ERG could not replicate the model proportions reported by the company.

The ERG agrees that, in the model, it is the proportion of patients in each KCCQ-CSS score that drives economic results, as KCCQ-CSS scores are not included in any capacity in the analysis. Nonetheless, the ERG notes that it remains crucial that the changes in KCCQ-CSS predicted by the model are validated against the observed data in EMPEROR-R.

The ERG is still unsure which dataset from EMPEROR-R was used to derive KCCQ-CSS changes in the model, although it believes that the results were based on the randomised set in EMPEROR-R with observed cases, including data after treatment discontinuation, with no imputation for deaths (Figure 20 of CS and Table 15.2.3.6:5 of the CSR). The ERG asks the company to confirm this.

These results show a change from baseline to week 12 of [redacted]; baseline to week 32 of [redacted]; and baseline to week [redacted], for placebo and empagliflozin, respectively, with baseline KCCQ-CSS scores of [redacted]. The baseline KCCQ-CSS scores would place patients in the model in KCCQ-CSS quartile 2 ([redacted]). Crucially, according to the observed changes in mean KCCQ-CSS in EMPEROR-R, SoC patients would only change KCCQ-CSS quartile in the model at week 52, while empagliflozin patients would only change KCCQ-CSS quartile from baseline to week 12 in the model.

Table 38 shows that in the model there were improvements in patients' KCCQ-CSS quartiles at all time points in both treatment arms. Importantly, the TPs used in month 9+ of the company's model assume that patients have a very small probability of leaving the KCCQ-CSS state they are in at month 8 in the analysis. This translates into a very strong assumption that the changes seen in EMPEROR-R from baseline to week 52 are sustained for approximately 30 years in the model and that the effect of empagliflozin lasts even after treatment discontinuation, as these patients never catch up to SoC patients.

Given that the there is a higher percentage of empagliflozin patients in the highest KCCQ-CSS state in the model at month 8, this difference is broadly maintained for the rest of the model time frame. Therefore, the company's assumption that empagliflozin patients experience SoC TPs after discontinuation is only partially conservative and leads to a sustained relative treatment effect for patients in KCCQ-CSS 4 in the model over time.

Due to the company's model structure, this also impacts the benefits associated with empagliflozin on HHF and mortality, as these outcomes are dependent on patients' distribution across KCCQ-CSS states.

Therefore, the ERG considers that further validation is required on the estimation of empagliflozin's relative effect on KCCQ-CSS. More specifically, the company should:

- Clarify which dataset from EMPEROR-R is being used to estimate the TPs;
- Provide the data from EMPEROR-R that allowed the estimation of TPs and proportion of patients in each KCCQ-CSS in the model;
- Produce the TPs observed in EMPEROR-R for the KCCQ-CSS quartiles defined in the model and explain how these relate to the mean changes reported in the trial;
- Conduct scenario analyses where the effect of empagliflozin seen at month 8 in the model (sustained by the combination of the proportion of patients in the better KCCQ-CSS states in the empagliflozin arm at month 8 and the low probability of disease progression for both SoC and empagliflozin arms in month 9+) wanes over time.

Table 38. Monthly KCCQ-CSS transition matrix

KCCQ-CSS transitions [From, To]	Empagliflozin + SoC			Placebo + SoC		
	Months 1-3	Months 4-8	Months 9+	Months 1-3	Months 4-8	Months 9+
KCCQ [1,1]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [1,2]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [1,3]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [1,4]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [2,1]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [2,2]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [2,3]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [2,4]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [3,1]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [3,2]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [3,3]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [3,4]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [4,1]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [4,2]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [4,3]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ [4,4]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

4.1.6.3 Hospitalisations for heart failure

The company used count data from EMPEROR-R trial to model the number of HHF in the model. There were 553 HHF events (including repeated hospitalisations) observed in the placebo arm and 388 events in the empagliflozin arm. The observed monthly rates of HHF in EMPEROR-R are reported in Figure 14. The company considered the rate of HHF to be constant over time in the two arms, with fluctuations starting after 20 months disregarded, given the company's assessment that the sample size was substantially reduced at that point.

Figure 14. Observed monthly HHF in EMPEROR-R (reproduced from Figure 10, Evidera 2021 appendix).

The monthly rate of HHF events was modelled using a Poisson model with generalised estimating equations with an auto-regressive covariance structure to account for correlations between repeated measures. The company also considered fitting the HHF data with a negative binomial distribution, however, concluded that the fitting procedure failed when using this distribution.

The final Poisson regression used in the model included two predictors: the time-varying KCCQ-CSS health state; and treatment received. The regression model was fitted to the ITT population (Table 39) in EMPEROR-R. The company also fitted the final regression model to different subgroups of the trial population, such as patients above and below 65 years (Table 40).

Table 39. Poisson regression for HHF, ITT population from EMPEROR-R

Covariate	Coefficient	SE	p-value
Intercept	[redacted]	[redacted]	[redacted]
Treatment effect Empagliflozin 10 mg	[redacted]	[redacted]	[redacted]
KCCQ-CSS: [redacted] (Quartile 2)	[redacted]	[redacted]	[redacted]
KCCQ-CSS: [redacted] (Quartile 3)	[redacted]	[redacted]	[redacted]
KCCQ-CSS: [redacted] (Quartile 4)	[redacted]	[redacted]	[redacted]

Table 40. Poisson regression for HHF, age subgroups from EMPEROR-R

Age ≥65				Age <65		
Covariate	Coefficient	SE	p-value	Coefficient	SE	p-value
Intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Treatment effect Empagliflozin 10 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ-CSS: [redacted] (Quartile 2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ-CSS: [redacted] (Quartile 3)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ-CSS: [redacted] (Quartile 4)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

4.1.6.4 ERG critique

Trial population

During clarification, the ERG noted its disagreement with the company's assessment that the rates of HHF in EMPEROR-R were constant over time. Figure 14 shows that the rates of HHF in the placebo arm are at least the same as those in the empagliflozin arm at around month 12; month 18; month 22; and month 27. While for some of these time points the numbers at risk are low, there are others where the number of patients at risk is still relatively high. According to Figure 15.2.2.1.2:2 in the CSR for EMPEROR-R, at 12 months (when the HHF rate appears to be the same in both treatment

arms) there were [redacted] of patients at risk of HHF in the placebo and empagliflozin arms, corresponding to [redacted] and [redacted] patients, respectively. At around 21 months, when the HHF rates begin to converge again (and eventually the HHF for placebo becomes lower than for empagliflozin), there were still approximately [redacted] of patients at risk in EMPEROR-R. Based on this, the ERG asked that the company considered an alternative approach to modelling HHF, allowing for the rate of HHF to vary over time within treatment arms.

The company replied that even though the regression model used to estimate HHF does not explicitly include a predictor for time, it includes a time-varying KCCQ-CSS predictor. The company also stated that the variation in Figure 14 may not necessarily indicate a real change in the rate between the arms. The company noted that the curves for placebo and empagliflozin at month 12 crossed but that the placebo rate increased sharply in month 13, exceeding the rate observed in immediately previous months and separating the curves again. The company concluded that the convergence at month 12 was, "*likely spurious and may not indicate a real change in the risk*". The company added that, "*the drop in HHF rate in the later part of the curves and their crossing may be due to high variability expected with low patient counts rather than a true signal of a change*."

The company conducted additional analysis to add time as a predictor in the regression model, as well as an interaction term between time and treatment. Coefficients (reported in Table 6 of the company's response to question B8) showed a negative slope for the time predictor with a p-value above 0.10. The company concluded that the declining rate of hospitalisation over time (suggested by the negative slope) was not clinically plausible. Therefore, rather than adding alternative equations to the model where the rate of HHF varied over time, the company included an option in the model to allow the user to turn off the treatment effect for empagliflozin on HHF at any chosen time.

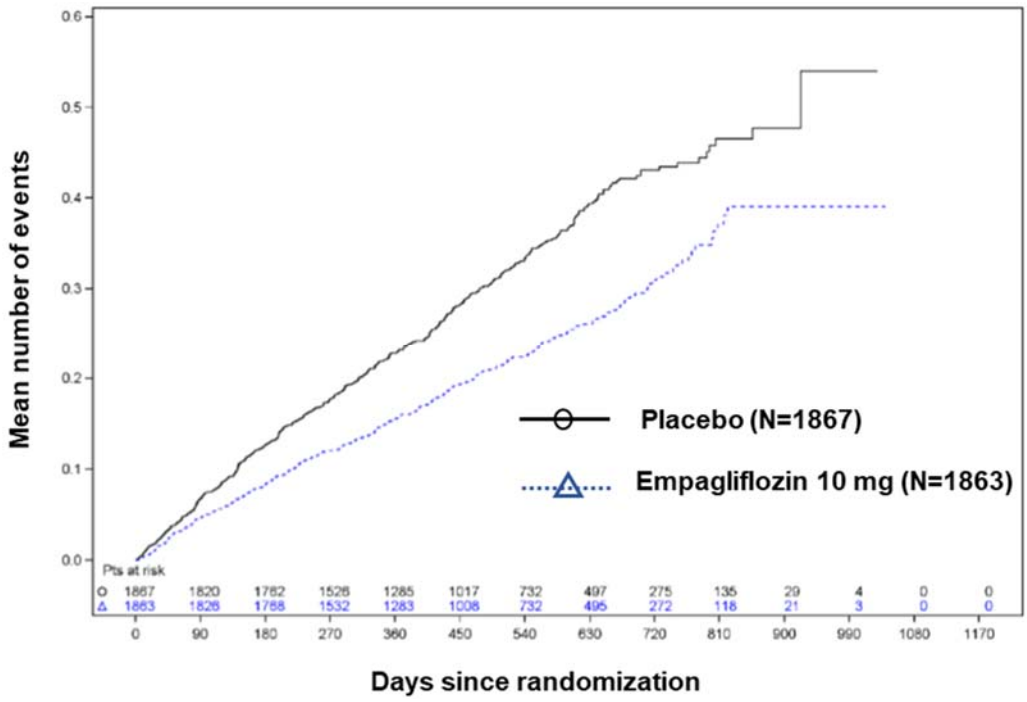
Even though the ERG agrees with the company that the time-varying KCCQ-CSS predictor in the HHF regression allows for HHF to vary over time in the model; and that a declining rate of HHF over time is not clinically plausible, the ERG notes that the underlying rate of HHF in the Poisson model is still constant. For example, if patients did not move KCCQ-CSS states in the model, the HHF would be the same in every model cycle for each treatment arm. Importantly, there is a lack of robust evidence to suggest that the overall rate for HHF remains constant over time across HFrEF patients.

The ERG also notes that that the option in the model to turn off the treatment effect for empagliflozin only addresses the company's assumption of a constant treatment effect across arms over time, and not the assumption of a constant HHF rate.

As a response to clarification question B9, the company provided the number of HHF predicted in the economic model, compared to the number of HHF in EMPEROR-R for the corresponding time period. The company's response compared the number of first HHF in EMPEROR-R (342 with placebo and 246 events with empagliflozin) with the number of HHF in the economic model for SoC and empagliflozin ([redacted] and [redacted], respectively). The ERG notes the following:

- The Poisson regression used to estimate HHF in the model included repeated hospitalisations for the same patient, therefore, the more appropriate number of events to include from EMPEROR-R would have been the total number of events - 553 and 388 events for the placebo and empagliflozin arms, respectively.
- The company's choice of time frame to calculate the number of predicted events by the economic model: the company calculated how many events happened in the first 16 months in the empagliflozin arm of the model, and in the first 15 months of the SoC arm of the model. The company reported that the median follow-up period of EMPEROR-R was 16 months, and this was the basis for the time frame chosen for the analysis. The ERG is unclear why a different number of months was chosen for the empagliflozin and the SoC arms of the model, and notes that this should have consisted of 16 months for both arms. The ERG notes that the mean follow-up period from EMPEROR-R is a more relevant time period for the analysis, however, given that median and mean follow-up period were the same in the trial, 16 months is appropriate for the analysis.
- When the ERG calculated the number of HHF in the model over 16 months (for the equivalent number of patients included in EMPEROR-R), it arrived at [redacted] HHF events estimated for the SoC and the empagliflozin arms, respectively. When compared to the observed 553 and 388 events of HHF for the placebo and empagliflozin arms of EMPEROR-R over the same period, the model demonstrates external face validity.

Figure 15. Mean cumulative function for occurrence of adjudicated HHF (first and recurrent) in EMPEROR-R (reproduced from Figure 8 CS)



Despite the model's ability to accurately reproduce the number of HHF observed in the trial, the ERG remains uncertain if HHFs are accurately estimated in the long-term for the trial population, given the company's assumption that HHF is constant and the use of the Poisson model.

Given that time to HHF Kaplan-Meier (KM) data were available from EMPEROR-R, the ERG considers that the company could have used these data to model time to HHF. Using the KM HHF data from EMPEROR-R would have allowed the company to fit a parametric survival curve to the data and extrapolate into the model's time horizon without having to assume a constant rate of HHF and without having to assume a constant treatment effect with empagliflozin. The ERG notes that using survival curves would have still allowed the company to model HHF by KCCQ-CSS state (as was done by the company for mortality – see Section 4.1.6.9).

Furthermore, using KM data for time to HHF would have allowed the company to model time to first and subsequent HHF separately. This could be of importance given the results reported in the EMPEROR-R CSR, indicating that time to subsequent HHF [redacted] in the empagliflozin than in the placebo arm (at 2 years, [redacted] of patients in the empagliflozin arm had experienced a second HHF, while [redacted] of patients had experienced a second event in the placebo arm).

Therefore, the ERG considers that the company's approach to modelling HHF could have relied on more appropriate methods.

Age subgroups from the trial

As discussed in Section 3.3.10.1, the treatment effect observed in patients in the <65 years category is likely to be driving the overall HHF treatment effect seen in EMPEROR-R, with hazard ratios (HRs) in the 65-75 years and ≥75 years categories being considerably higher than the HR for the <65 group and also non-statistically significant. This is also reflected in the Poisson coefficients reported in Table 40, where the treatment effect predictor becomes non statistically significant in the ≥65 years category. Therefore, the ERG is concerned that the size of the benefit associated with HHF for empagliflozin could be considerably smaller in an older population.

Patients with HFrEF in the UK

As discussed in Section 4.1.2, the company conducted a scenario analysis where the subgroup data from EMPEROR-R for patients above 65 years were used in the model to try and reflect the lower rates of HHF seen in PULSE and in clinical practice. Nonetheless, the ERG remains concerned that the company has not provided a scenario analysis which reflects the population seen in clinical practice. The analysis conducted by the company still severely overestimates HHF in the model when compared to PULSE (the best available estimate of the UK population).

In PULSE, there were 16,033 events observed for the 68,780 HFrEF patients over the maximum 5-year follow-up period in the study. The HHF estimates provided in PULSE are based on the number of person-years (p/y) reported in the study. This amounted to 204,862 p/y for the HFrEF population, which included a total of 68,780 patients. Therefore, the estimates provided are for a period of 3 years, which corresponds to the mean follow-up period in PULSE for HFrEF patients. The ERG compared these estimates from PULSE to the 3-year HHF outcomes in the model. As reported in Table 41, when compared to the PULSE results, the SoC arm of the model overestimates the number of HHF by more than double, which reinforces the model's lack of external validity when trying to estimate the impact of empagliflozin in the UK population.

Using the KM HHF data from the EMPEROR-R subgroup would likely still result in an overestimation of HHF when compared to the PULSE population, given the trial inclusion of sicker patients. However, adjusting HHF events in the model to reflect lower hospitalisations in both treatment arms would potentially be easier through the use of extrapolated curves, as these could be adjusted. For example, as it would appear that the HHF rate in EMPEROR-R is at least double that in PULSE,

applying a HR 0.5 to the extrapolated curves for empagliflozin and SoC would result in a more clinically plausible HHF rate for the HF population seen in clinical practice.

Table 41. Number of HHF events in PULSE and in the company's model for the ≥65 years population

Outcome	PULSE (N=68,780)	SoC arm of the model with ≥65 years population estimated for 68,780 patients	Difference
HHF events over 3-year follow-up in PULSE	16,033	[redacted]	[redacted]

4.1.6.5 Composite renal events

In every cycle of the model patients were at risk of experiencing a renal event. The company defined a renal event as the composite outcome including chronic dialysis, renal transplantation or a sustained eGFR reduction of ≥40% from baseline. Once patients experienced a renal event, a related cost and disutility associated with the renal event accrued every cycle of the model while patients were alive (discussed in Section 4.2.9 and Section 4.2.8, respectively). The monthly probability of patients experiencing a renal outcome while receiving treatment with empagliflozin was [redacted] compared to [redacted] for SoC. Both probabilities were directly taken from EMPEROR-R data.

Given that no patients experienced a renal transplantation in EMPEROR-R, the company assumed that no patients in the model underwent a transplant, therefore the events experienced in the renal event state consisted only of chronic dialysis or a sustained eGFR reduction of ≥40% from baseline. The company assumed that experiencing a renal event did not impact patients' survival in the model.

4.1.6.6 Treatment discontinuation

In order to estimate time to treatment discontinuation (TTD) in the model, the company fitted parametric survival curves to the TTD KM data from EMPEROR-R. The Weibull, Gompertz, log-logistic, lognormal, and the generalised gamma distributions were fitted to the TTD KM data and assessed for best visual fit; Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) statistics; and clinical plausibility.

Similar to HHF outcomes, the company included time-varying KCCQ-CSS and treatment predictors to the selected parametric models. Equations were also fitted with alternative parametric distributions for sensitivity analysis.

The company fitted a joint exponential model to both arms of the TTD data in EMPEROR-R, and it acknowledged that although it produced a poor fit based on AIC and BIC statistics and also a poor visual fit for the initial period of the KM TTD curves, it yielded the more plausible extrapolated mean TTD of approximately seven years. After treatment discontinuation with empagliflozin, patients were assumed to receive SoC only.

4.1.6.7 ERG critique

During clarification the ERG asked the company to only use the empagliflozin TTD KM data with an appropriate extrapolation in order to estimate empagliflozin costs in the model. The company agreed with the ERG's request and changed its base case.

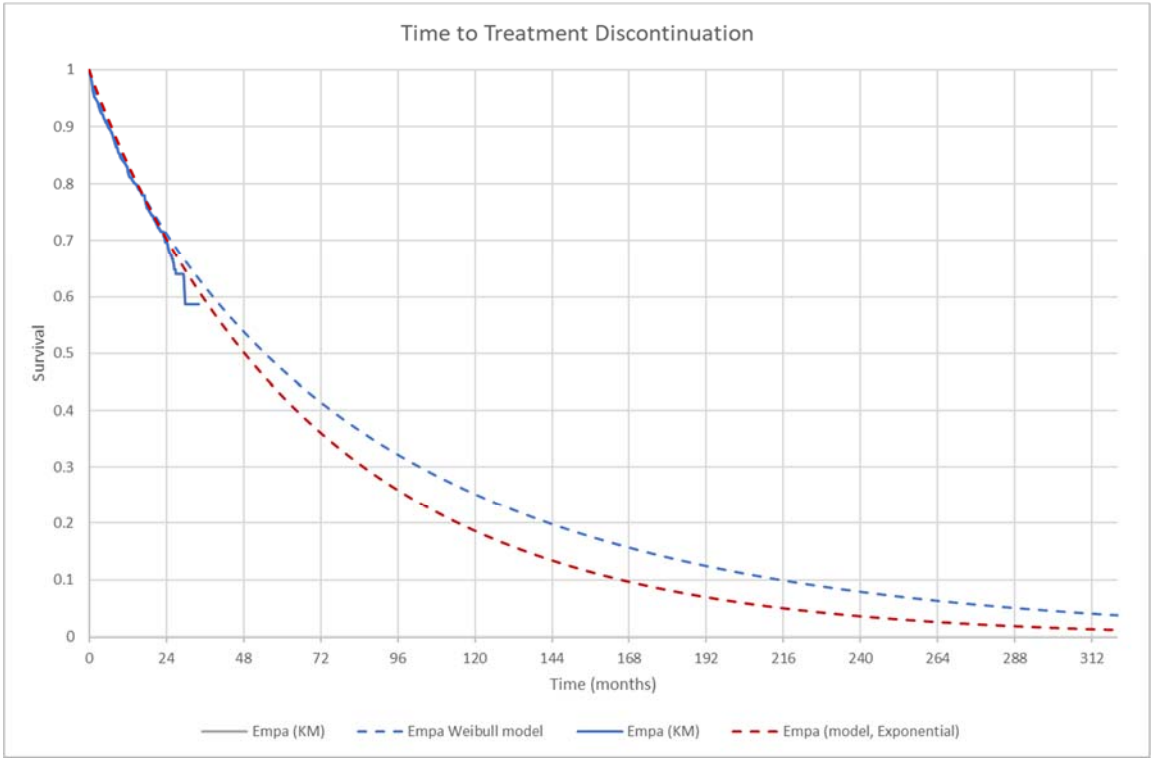
The CSR for EMPEROR-R reports that out of the 1,863 patients treated with empagliflozin, only 16% discontinued treatment for reasons other than death. When fatal events were included as a reason for discontinuation, the total percentage of patients discontinuing treatment amounted to 26%.

Expert opinion provided to the ERG was that lifelong treatment is expected for the majority of patients with HFrEF. Therefore, the ERG has conducted a scenario analysis where 84% of patients receive lifelong treatment with empagliflozin (and therefore, assumed the same probability of discontinuation in the model as that observed in EMPEROR-R). This assumption was corroborated by the ERG's clinical experts. In the dapagliflozin STA, an annual probability of discontinuation of 0.07 was assumed, which amounts to a similar probability to the 16% observed over 3 years in EMPEROR-R.

The ERG asked the company to justify the choice of the exponential model given its implicit assumption of a long-term constant rate of TTD. The company replied that there are no long-term TTD data for empagliflozin or dapagliflozin in HF to justify the longer-term assumptions. The ERG notes that the Weibull curve was the best fitting model according to AIC and Bic criteria, while the exponential provided the worse statistical fit. As acknowledged by the company, the Weibull curve provided a better visual fit to the beginning of the KM TTD curve but a worse fit to the end of the KM TTD curve when compared to the Weibull curve (Figure 16). The ERG agrees that neither curves provide an ideal visual fit to the tail of the KM TTD curve. However, given this is the most unreliable

portion of the TTD data; the better statistical fit of the Weibull curve; and the less strong assumptions of the Weibull, the ERG conducted a scenario analysis where the Weibull curve was used to model TTD.

Figure 16. Fitted curves to TTD KM data



4.1.6.8 Mortality

The company fitted parametric survival curves to the overall survival (OS) and to CV-related mortality KM data from EMPEROR-R. The Weibull, Gompertz, log-logistic, lognormal, and the generalised gamma distributions were fitted to the data and assessed for best visual fit; AIC and BIC statistics; and clinical plausibility.

Similar to TTD outcomes, the company included time-varying KCCQ-CSS and treatment predictors to the selected parametric models for each survival function. Equations were also fitted with alternative parametric distributions for sensitivity analysis.

The economic model uses different probabilities for CV-related and non-CV related deaths (which differed by KCCQ-CSS and treatment arm). In order to estimate the probability of patients having a

non-CV death the company subtracted the monthly probability of CV-death from the probability of all-cause death in each cycle of the model.

To ensure that the probability of non-CV death was never higher than the corresponding age-and sex-specific probability for the general UK population (taken from the national UK life tables for 2020), the maximum probability of non-CV death between EMPEROR-R and the life tables was taken in every cycle of the model. UK life tables were adjusted to exclude CV-related deaths to avoid double counting.[39]

4.1.6.8.1 Overall survival

A total of 249 (13.4%) patients died in the empagliflozin arm of EMPREROR-R, while 266 (14.2%) patients died in the placebo arm. The KM OS data reported in

Figure 17 show a small separation between the two arms from around month 17 to month 27 when the curves join again (HR 0.92; 95% CI: 0.77 to 1.10; p-value 0.35). The company attributed the trajectory for the OS curves to the small treatment effect and noise from variations in timing and frequency of deaths rather than changes in the magnitude of the effect of treatment. The company also noted that at the end of the KM curves (where there were only 200 patients at risk) there were approximately three months (between months 23 and 26) where no deaths were observed in the placebo arm, which created a small plateau in the KM curve, ultimately leading to the crossing on KM curves.

Figure 17. Observed OS data in EMPEROR-R (reproduced from Figure 2, Evidera statistical appendix).

The company fitted a joint Weibull model to the empagliflozin and SoC arms of the EMPEROR-R OS data and reported that fitting the data independently to the treatment arms led to estimates of longer survival for SoC when compared to empagliflozin (due to the crossing in the KM curves). The Weibull model had the best AIC and BIC statistic (CS Table 37, page 137) and was considered to produce the most clinically plausible extrapolated tails (Figure 18).

Figure 18. Weibull model fitted jointly to EMPEROR-R OS KM data.

4.1.6.8.2 CV-related death

A total of 187 (10%) patients had a CV-related death in the empagliflozin arm of EMPEROR-R, while 202 (10.8%) patients experienced a CV-related death in the placebo arm. Overall, CV-related deaths represented 75.5% of all deaths in EMPEROR-R; thus, time to CV-related survival KM curves (Figure 19) are very similar to the OS curves in

Figure 17. Similar to OS curves, the company attributed the trajectory of the CV-related survival curves to the small treatment effect and noise from variations in timing and frequency of deaths rather than changes in the magnitude of the effect of treatment.

The company also fitted a joint Weibull model to the empagliflozin and SoC arms of EMPEROR-R data. The company reported that fitting the data independently to the treatment arms led to estimates of longer survival for SoC when compared to empagliflozin and produced log-log survival vs log time curves for CV mortality (Figure 7 in the Evidera statistical appendix). The company concluded that, "*while the curves are not exactly parallel for the two treatment arms, the deviations are largely due to the crossing of the curves in the tail, which is likely caused by small patient counts*."

The Weibull (together with the log-logistic) model had the best AIC and BIC statistic (CS Table 37, page 137) and was considered to produce the most clinically plausible extrapolated tails, therefore was used in the company's base case analysis.

Figure 19. Observed CV-related mortality data in EMPEROR-R (reproduced from Figure 6, Evidera statistical appendix).

The company also fitted the final regression model to different subgroups of the EMPEROR-R trial population, such as patients above and below 65 years (Table 42).

Table 42. Weibull model for CV deaths, age subgroups from EMPEROR-R

Age ≥65				Age <65		
Covariate	**Coefficient**	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**
Shape	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Scale	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Treatment effect Empagliflozin 10 mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ-CSS: [redacted] (Quartile 2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ-CSS: [redacted] (Quartile 3)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ-CSS: [redacted] (Quartile 4)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

4.1.6.8.3 ERG critique

The company's approach to modelling all-cause and CV-related mortality assumes the existence of proportional hazards (PH) between the empagliflozin and SoC mortality curves.

The ERG considers that the assumption of a constant treatment effect (and therefore PHs) of empagliflozin over SoC throughout the model is unsubstantiated, both for all-cause and CV-related death. The ERG notes that the empagliflozin and placebo KM OS curves hardly separate during the follow-up period of the trial and that the HRs in EMPEROR-R for all-cause and CV-related death were not statistically significant (and also signaled a small effect size). Furthermore, it is the ERG's opinion that the log-log survival vs log time curves for CV mortality (Figure 7 in the Evidera statistical appendix) confirm that PHs are unlikely to hold for CV mortality.

As discussed in Section 4.1.2, the clinical experts advising the ERG noted that the tails of the Weibull distribution used in the company's base case was more representative of patients seen in the UK clinical practice, while the long-term predictions of the Gompertz curves would be a better representation of the higher mortality expected in the trial population.

Trial population

Given the company's own assessment that there were not enough CV deaths in EMPEROR-R to establish a robust effect of empagliflozin, and that CV-related deaths represented 75.5% of all

deaths in EMPEROR-R; the ERG considers that there is not enough evidence to support the inclusion of a treatment effect in non-CV mortality in the economic model. Furthermore, the company has not provided any clinical justification as to why non-CV related mortality in EMPEROR-R was expected to differ across treatment arms. The majority of non-CV related deaths (64 events in placebo and 62 in empagliflozin) were due to infection (21 events in placebo and 23 in empagliflozin), followed by malignancy (14 in placebo and 8 in empagliflozin).

The ERG notes that the Weibull curves fitted to the OS KM placebo and empagliflozin arms of the trial do not provide a bad fit up to month 27 (seen in Figure 18). However, there is no evidence to substantiate the separation in the curves modelled by the company from month 27 onwards. Therefore, the ERG conducted a scenario analysis where non-CV mortality was assumed the same in both treatment arms. As discussed above, the analysis conducted by the ERG also used the Gompertz distribution. Results for the scenario analysis conducted by the ERG are presented in Section 6.

During clarification, the ERG also requested that the company provided independently fitted curves to the trial arms in order to extrapolate CV-related deaths in the model. The ERG was concerned that the fitted Weibull model was a poor fit to the underlying CV mortality KM data, particularly to the empagliflozin arm (Figure 20) and therefore, recommended that the company provided a more flexible modelling approach.

The company did not undertake such analysis because it considered that independently fitted curves led to implausible long-term patterns with projections for placebo leading to longer mean time to CV-death with all the tested distributions. The company added that CV deaths were rare events in EMPEROR-R .

The ERG remains concerned that the jointly fitted Weibull curves are overestimating the treatment effect of empagliflozin on preventing CV-related deaths. Figure 20 shows that the fitted Weibul curves are a poor fit to the underlying KM data, with a constant overestimation of the CV survival curve for empagaflozin.

For the same reasons discussed above for all-cause mortality, the ERG also considers that there is not enough evidence to support the inclusion of a treatment effect for empagliflozin on CV mortality in the economic model. Further to this, the ERG notes the small size and the of lack of statistical significance of the treatment predictor in the CV mortality Weibull regression (Table 38 of CS;

coefficient for the treatment predictor on CV mortality of [redacted], p-value 0.562). Therefore, the ERG conducted a scenario analysis where no treatment effect was assumed for empagliflozin.

It is important to note that when no treatment effect is assumed for empagliflozin on mortality in the economic model, there is still a benefit associated with empagliflozin on both CV and non-CV mortality. This is because the coefficients for the KCCQ-CSS predictors (shown in Table 38 of the CS) differ according to health state, meaning that the probability of patients dying is different in every KCCQ-CSS state of the model. Given that patients in the empagliflozin arm of the model have a higher probability of remaining in the better KQCC-CSS states over time compared with SoC patients, the former also experience a lower probability of death.

This is a direct consequence of the lack of flexibility in the economic model, which structure intrinsically links the key clinical outcomes to the distribution of patients across the KCCQ-CSS states.

Figure 20. Weibull model fitted jointly to EMPEROR-R CV-related mortality KM data (taken from CS Figure 22).

Patients with HFrEF in the UK

During clarification, the ERG requested that the company ran the economic model on an older, less severe population, reflective of UK patients with HFrEF (with less HHF and longer survival). The company conducted a scenario analysis where the data for the subgroup analysis from EMPEROR-R for patients above 65 years in the trial were used. The company has not provided the KM data for all-cause or CV mortality in the above 65 years group, therefore the ERG cannot validate the newly fitted Weibull regressions against the appropriate KM data from the EMPEROR-R subgroup.

As reported in Table 43, when compared to the PULSE results, the SoC arm of the model overestimates the number of CV-related deaths (and underestimates the number of non-CV deaths) when the subgroup from EMPEROR-R is modelled. This reflects a population more likely to die of CV causes than PULSE patients.

Similar to HHF estimates, the ERG notes that the death estimates from PULSE are based on the number of person-years (p/y) reported in the study. This amounted to 204,862 p/y for the HFrEF population, which included a total of 68,780 patients. Therefore, the estimates provided are for a period of 3 years, which corresponds to the mean follow-up period in PULSE for HFrEF patients. The ERG compared these estimates from PULSE to the 3-year HHF outcomes in the model.

The ERG remains concerned that the results of the company's analysis on the cost-effectiveness of empagliflozin in the UK population with HFrEF are not reliable as these do not reflect the HHF or mortality (especially in the short-term) expected in the whole population considered in this appraisal. The lack of flexibility in the company's model structure and the lack of KCCQ-CSS data from PULSE mean that the company cannot use mortality data from PULSE directly in the SoC arm of the model. However, given the availability of KM CV and non-CV mortality data from PULSE, the ERG considers that this could be done by adjusting the KM curves from EMPEROR-R to more closely reflect the mortality curves in PULSE.

The ERG also conducted scenario analyses where no treatment effect was assumed for empagliflozin on mortality for this subgroup. Similar to the ITT trial population, when no treatment effect is assumed for empagliflozin in the economic model, there is still a benefit associated with empagliflozin on both CV and non-CV mortality.

Table 43. Number of deaths in PULSE and in the company's model for the ≥65 years population

Outcome	PULSE (N=68,780)	SoC arm of the model with ≥65 years population estimated for 68,780 patients	Difference
CV-related deaths	7,905	[redacted]	[redacted]
Non-CV deaths	9,599	[redacted]	[redacted]
All-cause deaths	17,504	[redacted]	[redacted]

4.1.7 Adverse events

The company used the most common adverse events (AEs) of special interest observed in EMPEROR-R to estimate AEs in the model. The AE rates per 100 patient years (reported in Table 44 of the CS) included urinary tract infections, mycotic genital infections, acute renal failure, hepatic injury, volume depletion, hypotension, hypoglycaemic event, and bone fractures. Patients who discontinued treatment with empagliflozin, were assumed to experience the rate of AEs associated with SoC.

4.1.7.1.1 ERG Critique

The ERG notes that the company's conversion of AE rates into probabilities (to be used in the economic model) was not done correctly. The company divided the AE rates per 100 patient years by 100 and the by 12, in order to get the monthly estimate per patient. However, the correct conversion would have been to apply the [1-exp (- rate/100/12)] formula to derive a monthly probability. Given the low AE rates (and thus low probabilities), using the correct formula would have only yielded marginally different results. The ERG notes that all monthly probabilities used in the model were below 1% and therefore, applying the ERG correction in the model had a negligible impact on the final results.

4.1.8 Health-related quality of life

EQ-5D-5L data were collected in the EMPEROR-R trial at baseline, 12 weeks, 32 weeks, 52 weeks, 100 weeks, and 148 weeks following randomisation, as well as at treatment discontinuation. Patients were also followed up 30 days following completion of the treatment period. In line with NICE guidance, the company mapped the EQ-5D-5L responses onto the EQ-5D-3L value set using the van Hout *et al.* 2012 algorithm.

Data after week 100 were excluded from the analysis because the number of available observations beyond this point dropped substantially (32% reporting at week 100 and only 2% at week 148). All

patients in the ITT population who had utility measurements available at baseline and at least one other later date were included. No imputation was used for visits where EQ-5D data were missing.

Utility scores were analysed using mixed-effects linear regression using all available EQ-5D measurements across all visits. The final model incorporated time-varying predictors reflecting whether a patient had their first hospitalisation in the last 0-1 months, 1-2 months, 2-4 months, and 4-12 months, while also accounting for disease severity by including KCCQ-CSS state predictors. The reference group had no HHF events to date, and patients were classified back into the reference group once a year had passed from hospitalisation. The model also included predictors for AEs (urinary tract infections, mycotic genital infections, acute renal failure, hepatic injury, volume depletion, hypotension, hypoglycaemic event, and bone fractures). The same approach was taken for each AE predictors, where patients were returned to the reference group one month after experiencing the AE. Other predictors included in the final regression were sex, age, region, and history of heart failure (Table 44).

Table 44. Final QoL regression used in the model

Covariate	Coefficient	SE	p-value
Intercept	[redacted]	[redacted]	[redacted]
Demographics			
Sex: Male	[redacted]	[redacted]	[redacted]
Age ≥65 years	[redacted]	[redacted]	[redacted]
Region			
Region: Asia	[redacted]	[redacted]	[redacted]
Region: Latin America	[redacted]	[redacted]	[redacted]
Region: North America	[redacted]	[redacted]	[redacted]
Region: Other	[redacted]	[redacted]	[redacted]
KCCQ-CSS			
KCCQ-CSS: 55 to <75 (Quartile 2)	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 75 to <90 (Quartile 3)	[redacted]	[redacted]	[redacted]
KCCQ-CSS: 90 to 100 (Quartile 4)	[redacted]	[redacted]	[redacted]
Baseline EQ-5D (standardised)	[redacted]	[redacted]	[redacted]
Medical History			
HF: Ischaemic cause	[redacted]	[redacted]	[redacted]
Time Since HHF			

Covariate	Coefficient	SE	p-value
HHF: <1 month	[redacted]	[redacted]	[redacted]
HHF: 1 to <2 months	[redacted]	[redacted]	[redacted]
HHF: 2 to <4 months	[redacted]	[redacted]	[redacted]
HHF: 4 to <12 months	[redacted]	[redacted]	[redacted]
AEs			
Urinary tract infection	[redacted]	[redacted]	[redacted]
Genital mycotic infection	[redacted]	[redacted]	[redacted]
Acute renal failure	[redacted]	[redacted]	[redacted]
Hepatic injury	[redacted]	[redacted]	[redacted]
Volume depletion	[redacted]	[redacted]	[redacted]
Hypotension	[redacted]	[redacted]	[redacted]
Hypoglycaemic event	[redacted]	[redacted]	[redacted]
Bone fracture	[redacted]	[redacted]	[redacted]
*taken from Sullivan *et al.* 2006			

Unlike for HHF and survival outcomes, treatment was not included as a predictor in the final model as it was found to be not statistically significant.

The company's estimation of the utility values for the four KCCQ-CSS states did not make use of all the predictors included in the final regression model. In order to estimate KCCQ-CSS utility values, the company used the characteristics from EMPEROR-R and multiplied the mean values observed at baseline by the coefficients for sex; age; region; baseline EQ-5D; and history of heart failure (i.e., ischaemic event) from the regression model and added the intercept coefficient. To these values, the company added the coefficient for each respective KCCQ-CSS predictor and generated four utilities values. Time since HHF and AEs predictors were not used to estimate utility values, but instead were used to estimate separate disutility values.

The estimated utility value for quartile 4 (the least severe quartile) was higher ([redacted]) than the utility seen in the UK general population reported by Sullivan *et al.*[1] for the age group of 60-69 years. Therefore, the company used the utility reported in Sullivan (0.774) for quartile 4 and adjusted the utility values for the 1–3 quartile states based on the absolute difference between the estimated utility for quartile 4 and the value reported in Sullivan (resulting in a [redacted] difference estimated as [redacted] minus 0.7740). The utility values are presented in Table 45.

Table 45. Utility values used in the company's base case

KCCQ-CSS Quartile Health State	Value used
Quartile 1	[redacted]
Quartile 2	[redacted]
Quartile 3	[redacted]
Quartile 4	0.7740

The annual disutility associated with HHF in the model was estimated as [redacted] per event. This was calculated by multiplying the coefficients in Table 44 for time since HHF by the respective period of time and adding these together [redacted].

The model applies a disutility of [redacted] to patients who experience a composite renal event. The company calculated this as a weighted average of disutilities associated with dialysis and CKD stage 3 disease observed in EMPEROR-R. The disutilities were sources from Jesky *et al*.[40]

The disutility values used for genital mycotic infection, acute renal failure, hepatic injury, volume depletion and bone fracture were those of the coefficients reported in Table 44. The utility value for hypotension was assumed equal to that of essential hypertension and taken from Sullivan 2006. Even though the company's regression analysis captured urinary tract infection events, the disutility associated with this event was sourced from Sullivan 2016[35] (-0.025).

4.1.8.1 ERG critique

KCCQ-CSS state utilities

[redacted]

Table 46. Change in EQ-5D scores in EMPEROR-R

Time	Empagliflozin 10 mg		Placebo		Difference (SE; p-value)
	N	Mean (SD)	N	Mean (SD)	
Baseline EQ-5D-3L	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline at week 12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline at week 32	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Change from baseline at week 52	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline at week 100	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline at week 148*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
*Not included in analyses due to low patient counts.					

The majority of the QALY gain in the model comes from the additional time spent by empagliflozin patients in quartile 4 of the KCCQ-CSS state when compared to SoC patients. The ERG notes that the utility value associated with the latter (0.7740) was taken from Sullivan for the age group of 60-69 years, however, corresponded to a population composed of 52% females and 48% males. The ERG notes that in EMPEROR-R, only 24% of the population were females.

Furthermore, the ERG disagrees with the company's adjustment applied for the 1–3 KCCQ-CSS quartiles, given that it was based on the absolute difference between the EMPEROR-R utility and the Sullivan value. As a response to an ERG's request, the company conducted a scenario analysis where the relative difference ([redacted]) was used instead, resulting in a change in utility values from [redacted], 0.7740 for KCCQ-CSS Q1 to Q4, respectively, to [redacted], 0.7740.

During clarification the ERG also requested that the company included age-related utility decrements throughout the model time horizon using the algorithm published by Ara and Brazier 2010, which the company supplied as a scenario analysis.

Overall, the ERG is concerned that the EQ-5D data from EMPEROR-R [redacted] and that the economic model generates a QALY gain of [redacted]. The ERG notes that patients' QoL gain is related to how much longer patients stay in the better KCCQ-CSS states and as discussed in Section 4.1.6.2, the benefit associated with empagliflozin on KCCQ-CSS transitions is maintained after treatment discontinuation in the model.

The ERG is further concerned that the 0.7740 utility value does not accurately reflect the baseline gender distribution in EMPEROR-R and therefore, recommends that the company adjusts this value to reflect the baseline gender split in the trial.

Furthermore, the ERG considers that the relative utility adjustment and the age-related decrements from Ara should be use used in the base case results, and therefore reports the results of these analysis in Section 6.

The ERG also has issues with how the company conducted their scenario analysis using the older age subgroup from EMPEROR-R. The baseline characteristics from the older subgroup were used in the regression analysis reported in Table 44, however the regression was not re-estimated in this subgroup and thus the coefficients for the predictors remained the same. The ERG, therefore recommends that the company re-estimates the regression model using the subgroup data. Furthermore, the utility value for quartile 4 of the KCCQ-CSS state taken from Sullivan study for the age group of 60-69 years also remained unchanged. Nonetheless, the study provided a value of 0.723 for the age group of 70-79 years, which the company should have used instead. The ERG notes that the gender distribution in Sullivan population is still not fully representative of the that observed in UK clinical practice (around 35% females). Therefore, the ERG conducted an initial exploratory analysis where the 0.723 value was used (and all other KCCQ-CSS state values were adjusted accordingly) however, recommends that the company adjusts this value to better reflect the gender distribution in UK HFrEF patients.

Disutility associated with HHF

The second biggest driver of the QALY gains in the model comes from the reduction in HHF for empagliflozin patients when compared to SoC patients. The ERG considers that the impact of HHF on QoL is overestimated in the model. The company arrived at the [redacted] disutility per HHF event by multiplying the coefficients in Table 44 for time since HHF (<1 month; 1 to <2 months; 2 to <4 months; 4 to <12 months) by the respective period of time and adding these together [redacted]), therefore arriving at an annual disutility value. Nonetheless, the disutility was applied every month of the model, to the number of new monthly HHF events. Therefore, the company is estimating the annual impact that a hospitalisation lasting for 1 year has on patients' QoL.

The ERG notes that it did not have access to mean (or median) duration of hospitalisation in EMPEROR-R, and so it cannot ascertain the extent to the overestimation of this disutility. Nonetheless, the ERG does not consider it a reasonable assumption that all patients would be hospitalised for 1 full year.

Therefore, the ERG advises that the company reports the proportion of patients in EMPEROR-R who were hospitalised for 1; 2; and 8 months in the trial and generates a weighted disutility value to be applied in the model. Ideally, the same analysis would be conducted for PULSE data, as the mean duration of HHF is likely to be lower in PULSE than in EMPEROR-R.

Disutility associated with AEs

Even though the company used disutilty values estimated from its regression analysis that were not statistically significant, and despite some inconsistencies in the sources for the disutilities used, the scenario analysis conducted by the company during clarification (question B25 and B26) assured the ERG that the AE-related disutilities used in the model have a negligible impact on the final ICER.

4.1.9 Resource use and costs

4.1.9.1 Treatment and comparator costs

The intervention included in the economic model was empagliflozin formulated as a 10mg tablet taken once a day, in addition to SoC. The list price for empagliflozin is £36.59 for a pack of 28 pills, amounting to a daily cost of £1.31 and a monthly cost of £39.78.

Standard of care was modelled as a basket of drugs used in first-line heart failure care. The CS describes first line SoC as either an angiotensin-converting enzyme inhibitor (ACEi), or an angiotensin receptor-blocker (ARB), if patients are intolerant to ACEi, with a beta-blocker (BB). If symptoms continue, mineralocorticoid receptor antagonists (MRA) can be added to the ACEi or ARB if there is no evidence of hyperkalaemia. Some patients are also prescribed sacubitril valsartan, an ARNi (angiotensin receptor-neprilysin inhibitor). Patients are also routinely prescribed diuretics to provide symptomatic relief, particularly in the presence of oedema.

The proportion of drugs in the SoC basket was taken from the baseline distribution of treatments in EMPEROR-R . The proportions, unit costs and monthly costs are given in Table 47. Where there were multiple drugs in a class, an average of the unit costs was taken. The total monthly cost of SoC in the model is £44.22 and the total cost of empagliflozin + SoC is £84.00 per month.

All drugs prescribed in the intervention and comparator arms of the model are orally administered thus, attract no administration costs.

Table 47. SoC drug costs

Drug class	ITT Proportion	Monthly cost
ACEi	45.4%	£6.03
ARB	24.3%	£11.05
ARNi	19.5%	£119.44
MRA	71.3%	£7.63
Beta blocker	94.7%	£9.50
Loop or high ceiling diuretics	84.5%	£1.28

Abbreviations: ACEi. angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor-blocker; MRA, mineralocorticoid receptor antagonists

4.1.9.2 ERG critique

The clinical experts advising the ERG agreed with the drugs included in the company's basket of SoC treatments, however, disagreed with some of the proportions used. The clinical experts indicated that the proportion of UK patients who receive ACEi is higher than that assumed by the company (around 62%), while the proportion of patients receiving an ARNi (sacubitril valsartan) is lower and amounts to 10%. Therefore, the ERG conducted a scenario analysis to reflect the clinical expert opinion and presented the results in Section 6.

4.1.9.3 Disease management costs

Disease management costs included GP and cardiologist visits, and A&E referrals. Resource use was based on data from the Clinical Practice Research Datalink, as reported by McMurray *et al.* (2018)[33]. Unit costs were taken from PSSRU Unit Costs of Health and Social Care[41] and the Schedule of NHS Costs.[42] The company assumed that GP visits were based on patient contact lasting 9.22 minutes and that and cardiologist visits were consultant-led, face to face follow-up appointments. The cost of an A&E referral was a weighted mean derived from national average unit costs and number of finished consultant episodes (FCEs) for non-admitted emergency medicine. All disease management unit costs were inflated to 2021 by applying the consumer price health inflation factor from Eurostat[43], and are reported in Table 50 of the CS.

4.1.9.4 ERG critique

The ERG's clinical experts considered the GP rates and cardiologists to be high compared to what they would expect in UK clinical practice (Table 48). Therefore, during clarification the ERG

requested the company implement a scenario analysis to include the resource use based on the ERG's clinical experts' estimates. The impact on the final ICER was negligible.

Table 48. Disease management resource use

Resource	Annual visits (company's base-case)	ERG's clinical expert (per year)
GP visit	23.14	5 GP visits a year, and 18 nurse visits a year
Cardiologist visit	0.05	0.20
A&E referral	0.01	0.02

4.1.9.5 Hospitalisation costs

The acute costs of HHF were based on NHS reference costs[42] for non-elective long inpatient stay, computed as the weighted average of reference costs for healthcare resource group (HRG) codes (described in Table 49) and the number of FCEs. The total cost of HHF in the model was £3,072 (inflated to 2021 by applying the consumer price health inflation factor from Eurostat[43]).

Table 49. Hospitalisation for heart failure unit costs

Reference cost code and description	Unit cost	Number of events	Weighted cost
Heart Failure or Shock, with CC Score 14+ (EB03A) – Non-Elective (Long Stay)	£3,909.61	23,406	£932.55
Heart Failure or Shock, with CC Score 11-13 (EB03B) – Non-Elective (Long Stay)	£3,139.47	28,511	£912.18
Heart Failure or Shock, with CC Score 8-10 (EB03C) – Non-Elective (Long Stay)	£2,532.67	24,564	£634.00
Heart Failure or Shock, with CC Score 4-7 (EB03D) – Non-Elective (Long Stay)	£2,169.60	18,805	£415.78

Heart Failure or Shock, with CC Score 0-3 (EB03E) – Non-Elective (Long Stay)	£2,169.93	2,841	£62.82
Abbreviations: CC, critical care			

4.1.9.1 CV-related mortality

The cost of CV death was based on the regression analysis presented in Alva *et al.* which estimated the added inpatient costs for type-2 diabetes (T2D) complications, during the UK Prospective Diabetes Study post-trial monitoring period from 1997 to 2007, and used hospitalisation records for patients in England (n=2,791).[38]

The regression analysis reported coefficients for the expected cost impact of T2D complications, which included fatal myocardial infarction (MI), fatal ischaemic heart disease (IHD), and fatal stroke, as well as age and gender, on inpatient hospitalisation costs. The company used these coefficients to estimate the costs of fatal MI; IHD and stroke. Costs were estimated separately for males and females; aged <65 years and ≥65 years, respectively.

The company then used the percentage of males/females and the percentage aged <65 or ≥65 years from EMPEROR-R and derived a weighted average cost for each event. A simple average was then taken across the cost of the three fatal events to derive the final cost of CV death for the model of £4,146 (inflated to 2021 by applying the consumer price health inflation factor from Eurostat).[43]

These costs were applied as one-off events in the model. Non-CV deaths were assumed to incur no cost in the base case.

4.1.9.2 ERG critique

The ERG disagrees with the use of the chosen estimates from the Alva paper as these relate to the added costs on hospitalisations due to T2D complications. Therefore, during clarification the ERG asked that the company used the alternative estimates provided in Table 3 of the Alva study, which reported the absolute cost of the events. The company conducted this analysis and reported that the cost of CV death changed from £4,146 to £3,733 (after price inflation).

During clarification, the ERG also noted that [redacted] Therefore, instead of using a simple average across the cost of the three fatal events (MI; IHD and stroke), the ERG asked that the company

weighted the cost of a CV-death by the proportion of fatal CV events observed in EMPEROR-R, with the cost of sudden cardiac death as the highest contributor to the costs associated with CV deaths in the model.

To address the ERG's concerns the company conducted two analyses, both of which included the cost of sudden cardiac death as the highest contributor to the costs of CV deaths and reflected the distribution of fatal CV events observed in EMPEROR-R.

In the first analysis, the company assumed the cost of sudden cardiac death to be zero and in the second analysis, the company used a unit cost of £1,632 for all sudden cardiac deaths in the model corresponding to the total HRG costs for cardiac arrest (NHS Costs 2019-20). In both analyses, the total cost of CV-death was calculated as a weighted average of the relevant fatal events listed in the EMPEROR-R CSR.

With the ERG's preferred unit costs from Alva, the alternative total cost of CV deaths are: £3,733 (company's base case); £1,582 (with the cost of sudden death assumed to be zero); or £2,307 (with the cost of sudden death assumed to be £1,632).

The ERG's preferred cost for CV death in the model is £1,582 as it represents the most conservative estimate. The ERG notes that the cost of CV death accepted by the committee for the dapagliflozin STA was £1,739 (price uplifted to 2020). Scenario analysis including this estimate are reported in Section 6.

4.1.9.3 Composite renal outcome

The costs associated with the composite renal outcome were taken from published costs of the individual renal outcomes (i.e., dialysis and sustained eGFR reduction). The unit costs for dialysis were sourced from Kerr *et al.* 2012.[36]

To estimate the cost of sustained eGFR reduction, a 40% eGFR decline was applied to the mean eGFR at baseline in the EMPEROR-R trial (62.0 ml/min/1.73 m^2), resulting in an eGFR value of 37.2 ml/min/1.73 m^2 (i.e., CKD stage 3b). The unit cost for CKD stage 1-3B (£511.23, inflated to 2021 prices) was obtained from the study by Kent *et al*. 2015, who estimated annual UK hospital care costs by CKD stage. [37] The total cost for renal outcomes in the model was £4,862.

4.1.9.4 ERG critique

The ERG notes that the Kerr *et al.* 2012[36] study estimated the mean annual cost to the English NHS (2009-2010 prices) of direct care per patient, which the company inflated to 2021 costs. However, there is very little detail provided in the study for the resource use included in this estimate (frequency of dialysis, setting, patients' age etc.).

The ERG notes that the company could have used the 2019/2020 costs sourced from the NHS Cost Schedule[42], which report that the cost of hospital haemodialysis, with access via haemodialysis catheter (for 19 years and over) to be £148 per session (code LD01A). The ERG prefers the use of this unit cost given the lack of transparency around the estimate from Kerr *et al.* 2012[36]. Therefore, the ERG calculated the annual cost of dialysis (assuming 3 weekly sessions) as £23,088. Thus, the ERG preferred cost for renal outcomes amounted to £3,333. Results of the using this cost in the model are provided in Section 6.

4.1.9.5 Adverse event costs

The unit costs for outpatient visits were taken from the PSSRU unit costs of health and social care, while inpatient costs were taken from the National Schedule of NHS Costs for 2018/19 and then inflated to 2021 costs using the consumer price health inflation factor from Eurostat.[43]

4.1.9.6 ERG critique

Similar to the renal outcomes costs, the company could have used the 2019/2020 more up to date costs sourced from the NHS Cost Schedule, instead of using older cost estimates and inflating these.[42]

Furthermore, clinical experts advising the ERG noted that the split between outpatient and inpatient care for treating AEs in the company's model should have included a higher number of inpatient visits. The ERG ran a scenario analysis where the appropriate 2019/2020 costs in the most recent schedule were used, together with applying the ERG's clinical expert's distribution of inpatient/outpatient visits. Both had a negligible impact on the final ICER.

5 Cost effectiveness results

The results of the company's updated (i.e., post clarification) base case deterministic analysis are presented in Table 50. In the base case analysis, empagliflozin + SoC generates [redacted] incremental QALYs and incremental costs of £926 over SoC alone, resulting in an ICER of [redacted] per QALY gained.

Table 50. Company's base case results

Interventions	Total Costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
Empagliflozin	£17,837	[redacted]	[redacted]	£926	[redacted]	[redacted]	[redacted]
Standard of Care	£16,911	[redacted]	[redacted]	-	-	-	-

Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year;

The company performed probabilistic sensitivity analysis (PSA) to assess the joint parameter uncertainty around the base case results, using 1,000 PSA iterations. Table 51 shows that the company's PSA ICER of [redacted] per QALY gained is similar to the company's deterministic ICER. The probability of empagliflozin being cost effective at the £30,000 threshold is about 85% (as per Figure 21).

Table 51. Company's mean PSA results

Interventions	Total Costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
Empagliflozin	£17,719	[redacted]	[redacted]	£923.59	[redacted]	[redacted]	[redacted]
Standard of Care	£16,795	[redacted]	[redacted]	-	-	-	-

Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year;

Figure 21. Cost-effectiveness acceptability curve

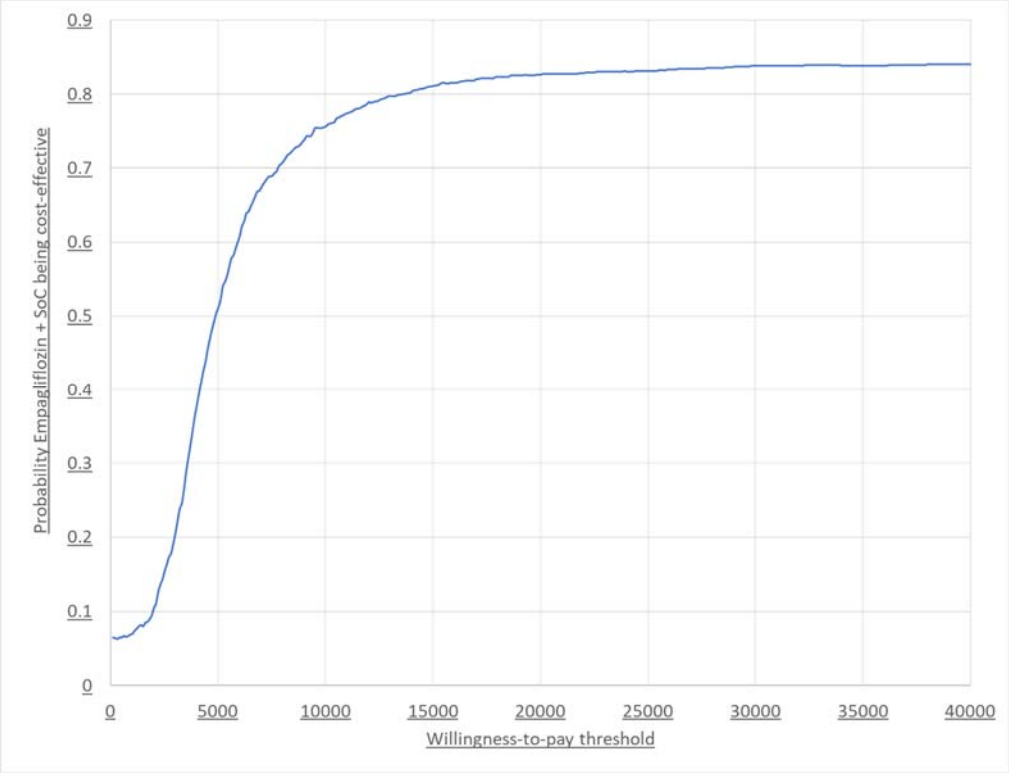


Figure 22. Cost-effectiveness plane

[redacted]

5.1.1 *Company's sensitivity analyses*

The company carried out one-way sensitivity analyses (OWSAs) to assess the impact of varying specific parameters in isolation on the final ICER and to identify the main model drivers. The company concluded that the main driver of results in the model was the treatment covariate included in the HHF regression equations. When this was excluded from the HHF estimations, the ICER increased to [redacted] Figure 28 CS).

The company also carried out scenario analyses changing assumptions surrounding key parameters, presented in Table 6 of the company's updated response to the ERG's clarification questions. The largest change in ICER occurred when renal outcomes were excluded from the model, which increased the ICER by 31% to [redacted] per QALYs gained.

6 Additional economic analysis undertaken by the ERG

6.1 Additional requests to the company

The ERG produced a list of issues requiring additional clarification or further analysis from the company. These have been discussed in detail throughout Section 4.

- KCCQ-CSS states:

1. The company should clarify which dataset from EMPEROR-R is being used to estimate the TPs;
2. The company should provide the data from EMPEROR-R that allowed the estimation of TPs and proportion of patients in each KCCQ-CSS in the model;
3. The company should produce the TPs observed in EMPEROR-R for the KCCQ-CSS quartiles defined in the model and explain how these relate to the mean changes reported in the trial;
4. The company should conduct scenario analyses where the effect of empagliflozin seen at month 8 in the model (sustained by the combination of the proportion of patients in the better KCCQ-CSS states in the empagliflozin arm at month 8 and the low probability of disease progression for both SoC and empagliflozin arms in month 9+) is waned over time.

-

- Hospitalisations:

5. The ERG recommends that the company undertakes a scenario analysis where HHF KM data from EMPEROR-R is used to model time to HHF in the trial population. Using the KM HHF data from EMPEROR-R would have allowed the company to fit a parametric survival curve to the data and extrapolate into the model's time horizon without having to assume a constant rate of HHF and without having to assume a constant treatment effect with empagliflozin.
6. The ERG recommends that the company undertakes a scenario analysis where HHF KM data from EMPEROR-R is used to model time to HHF in the UK population and adjusts extrapolated curves to reflect a lower number of total HHFs in the model.

-

- Mortality:

7. The company conducted a scenario analysis where the data for the subgroup analysis from EMPEROR-R for patients above 65 years in the trial were used. The company has not provided the KM data for all-cause or CV mortality in the above 65 years group, therefore the ERG could not validate the newly fitted Weibull regressions against the appropriate KM data from the EMPEROR-R subgroup. The ERG asks that the company supplies these.
8. Given the availability of KM CV and non-CV mortality data from PULSE, and the overestimation of mortality in the model for the UK population, the ERG recommends that the company adjusts the KM curves from the EMPEROR-R subgroup to more closely reflect the mortality curves in PULSE.

-
- Quality of life analysis:
  9. The ERG is concerned that the 0.7740 utility value does not accurately reflect the baseline gender distribution in EMPEROR-R and therefore, recommends that the company adjusts this value in the trial population analysis.
  10. The ERG is concerned that the 0.723 utility value does not accurately reflect the baseline gender distribution in the UK population and therefore, recommends that the company adjusts this value in the older subgroup analysis.
  11. The baseline characteristics from the older subgroup were used in the QoL regression analysis reported in Table 44, however the regression was not re-estimated in this subgroup and thus the coefficients for the predictors remained the same as those for the ITT population. The ERG, therefore recommends that the company re-estimates the regression model using the subgroup data.
  12. The ERG advises that the company reports the proportion of patients in EMPEROR-R who were hospitalised for 1; 2; and 8 months in the trial and generates a weighted disutility value to be applied in the model. Ideally, the same analysis would be conducted for PULSE data, as the mean duration of HHF is likely to be lower in PULSE than in EMPEROR-R.

6.2 Exploratory and sensitivity analyses undertaken by the ERG

The scenario analyses undertaken by the ERG are explained throughout Section 4 of the report. The ERG conducted two sets of analysis, one using the entire trial population from EMPEROR-R and the other using the above 65 years subgroup from the trial. As discussed throughout the report, the

former analysis aims to estimate the cost-effectiveness of empagliflozin vs SoC in the trial population, while the latter analysis intends to ascertain the cost-effectiveness of empagliflozin vs SoC in the UK population with HFrEF. Nonetheless, the ERG remains concerned that using the older subgroup of patients from EMPEROR-R still reflects a sicker population than that seen in UK clinical practice, with higher mortality and HHFs.

Results of the exploratory analyses conducted using the trial population are reported in Table 14, while Table 15 reports the results in the UK population analysis. The following analyses were condcuted in both populations:

7. Assuming that 84% of patients receive lifelong treatment with empagliflozin;
8. Using a Weibull model to estimate TTD for empagliflozin;
9. Using the relative utility adjustment and the age-related decrements from Ara;
10. Replacing the proportion of UK patients who receive ACEi and ARNi in the model to reflect the ERG's clinical experts' opinion;
11. Using a unit cost for CV death in the model of £1,582;
12. Applying the ERG-calculated annual cost of dialysis (assuming 3 weekly sessions) of £23,088;

The analyses listed below are specific to each population:

Trial population

d. Assuming that non-CV and CV mortality is the same for empagliflozin and SoC and using a Gompertz curve to estimate mortality.

UK population

e. Assuming that non-CV and CV mortality is the same for empagliflozin and SoC and using a Weibull curve to estimate mortality.
f. Using the 0.723 utility value from Sullivan for the KCCQ-CSS quartile state (and adjusting other KCCQ-CSS state values accordingly).

Results in Table 14 and Table 15 show that the ICERs in both populations do not change by more than approximately £1,000 per QALY gained. This is a direct consequence of the lack of flexibility in the economic model and it demonstrates how the key clinical outcomes (such as mortality) are intrinsically linked to the distribution of patients across the KCCQ-CSS states.

The key driver of the model is likely to be the distribution of patients across the KCCQ-CSS states of the model around month 8, given how these are maintained in the long-term analysis. The second key driver of the model is likely to be the rate of hospitalisations assumed in the model.

Table 52. Results of ERG's exploratory analysis in the trial population

	Results per patient	Empagliflozin	SoC	Incremental value
0	Company's base case post clarification			
	Total costs	£17,837	£16,911	£926
	QALYs	3.76	3.56	0.20
	ICER (£/QALY)	-	-	£4,717
1	Assuming that 84% of patients receive lifelong treatment with empagliflozin			
	Total costs	£18,433	£16,911	£1,522
	QALYs	3.86	3.56	0.30
	ICER (£/QALY)	-	-	£5,008
2	Using a Weibull model to estimate TTD for empagliflozin			
	Total costs	£17,893	£16,911	£981
	QALYs	3.77	3.56	0.21
	ICER (£/QALY)	-	-	£4,763
a	Assuming that non-CV and CV mortality is the same for empagliflozin and SoC using a Gompertz curve			
	Total costs	£12,798	£12,194	£604
	QALYs	2.63	2.53	0.11
	ICER (£/QALY)	-	-	£5,712
3	Using the relative utility adjustment and the age-related decrements from Ara			
	Total costs	£17,837	£16,911	£926
	QALYs	3.68	3.49	0.19
	ICER (£/QALY)	--	-	£4,915
4	Replacing the proportion of UK patients who receive ACEi and ARNi in the model to reflect the ERG's clinical experts' opinion			
	Total costs	£17,115	£16,212	£903
	QALYs	3.76	3.56	0.20
	ICER (£/QALY)	-	-	£4,602
5	Using a unit cost for CV death in the model of £1,582			
	Total costs	£16,405	£15,452	£952
	QALYs	3.76	3.56	0.20
	ICER (£/QALY)	-	-	£4,853
6	Applying the ERG-calculated annual cost of dialysis (assuming 3 weekly sessions) of £23,088			
	Total costs	£17,576	£16,565	£1,011
	QALYs	3.76	3.56	0.20
	ICER (£/QALY)	-	-	£5,152

Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

Table 53. Results of ERG's exploratory analysis in the UK population

	Results per patient	Empagliflozin	SoC	Incremental value
0	Company's base case post clarification using ≥65 subgroup data from EMPEROR-R + using Weibull curve to estimate mortality			
	Total costs	£16,436	£15,198	£1,238
	QALYs	3.55	3.36	0.20
	ICER (£/QALY)	-	-	£6,342
1	Assuming that 84% of patients receive lifelong treatment with empagliflozin			
	Total costs	£17,134	£15,198	£1,937
	QALYs	3.64	3.36	0.29
	ICER (£/QALY)	-	-	£6,795
2	Using a Weibull model to estimate TTD for empagliflozin			
	Total costs	£16,502	£15,198	£1,305
	QALYs	3.56	3.36	0.20
	ICER (£/QALY)	-	-	£6,410
b	Assuming that non-CV and CV mortality is the same for empagliflozin and SoC using a Weibull curve and the above 65 subgroup population in the model			
	Total costs	£16,168	£15,198	£971
	QALYs	3.49	3.36	0.13
	ICER (£/QALY)	-	-	£7,270
3	Using the relative utility adjustment and the age-related decrements from Ara[44]			
	Total costs	£16,436	£15,198	£1,238
	QALYs	3.47	3.28	0.19
	ICER (£/QALY)	-	-	£6,641
c	Using the 0.723 utility value from Sullivan for the KCCQ-CSS quartile 4 state (and adjusting other KCCQ-CSS state values accordingly)			
	Total costs	£16,436	£15,198	£1,238
	QALYs	3.34	3.16	0.18
	ICER (£/QALY)	-	-	£6,758
4	Replacing the proportion of UK patients who receive ACEi and ARNi in the model to reflect the ERG's clinical experts' opinion			
	Total costs	£16,406	£15,169	£1,237
	QALYs	3.55	3.36	0.20
	ICER (£/QALY)	-	-	£6,336
5	Using a unit cost for CV death in the model of £1,582			
	Total costs	£15,092	£13,841	£1,251
	QALYs	3.55	3.36	0.20
	ICER (£/QALY)	-	-	£6,407
6	Applying the ERG-calculated annual cost of dialysis (assuming 3 weekly sessions) of £23,088			

	Total costs	£16,217	£14,907	£1,311
	QALYs	3.55	3.36	0.20
	ICER (£/QALY)			£6,713

Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

6.3 Conclusions of the cost effectiveness sections

Two key areas of uncertainty remain in the economic analysis: the long-term effect of empagliflozin on patients' change in KCCQ-CSS (in both the trial and in the UK population analyses sets) and the lack of representativeness of the subgroup data from EMPEROR-R when trying to replicate the UK population. The overestimation of CV mortality and the overestimation of HHFs in the model compared to the UK population when the ≥65 years subgroup is used in the model indicate the lack of external validity of the model results in this population.

The ERG conducted one additional scenario analysis to try and explore these areas of remaining uncertainty. The scenario assumed that empagliflozin had no effect on patients' transitioning through KCCQ-CSS states. This means that SoC and empagliflozin patients were distributed equally across the KCCQ-CSS states in the economic model. When this assumption was used in the economic model (in combination with the assumptions described in Section 6.2), the final ICER for the trial population was £15,725, while for the UK population was £31,936 per QALY gained. In both cases, the main driver of QALY gain in the model was the benefit of empagliflozin on HHF.

Given that for the UK population analysis, the SoC arm of the model overestimates the number of HHF by more than double when compared to PULSE, the ERG notes that the ICER of £31,936 is likely to increase substantially, if the number of HHF was reduced in the model. The ERG could not robustly adjust the HHF rate in the model as this would entail artificially manipulating the coefficients for the HHF regression to produce a lower number of HHFs in the model. The ERG, therefore, recommends that the company conducts this analysis.

The scenario analysis conducted by the ERG indicates that the ICER for empagliflozin compared to SoC is likely to remain under the £30,000 threshold in the trial population, even when it is assumed that empagliflozin has no effect on patients' movements through the KCCQ-CSS quartiles defined by the company. Nonetheless, the ERG remains concerned that the cost-effectiveness of empagliflozin compared to SoC in the UK population remains highly uncertain.

7 End of Life

The company has not made a case for the committee to consider empagliflozin as an end of life treatment and the Evidence Review Group (ERG) agrees with this assessment.

8 References

1. Sullivan PW, Slejko JF, Sculpher MJ, Ghushchyan V. Catalogue of EQ-5D scores for the United Kingdom. *Med Decis Making* 2011; **31**: 800-4.
2. National Institute for Health and Care Excellence (NICE). Final scope: Empagliflozin for treating chronic heart failure with reduced ejection fraction [ID3826], 2021. Available from: https://www.nice.org.uk/guidance/gid-ta10719/documents/final-scope. Date accessed: 11 AUG 2021.
3. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur J Heart Fail* 2016; **18**: 891-975.
4. National Institute for Health and Care Excellence. Chronic heart failure in adults: diagnosis and management NICE guideline [NG106]. 2018. Available from: https://www.nice.org.uk/guidance/ng106. Date accessed: 01 Sept 2021.
5. Boehringer Ingelheim. Incidence prevalence and resource utilisation of heart failure in England: a linked descriptive analysis of English primary and secondary care – PULSE study. [Data on File]. 2021.
6. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance: Dapagliflozin for treating chronic heart failure with reduced ejection fraction [TA679], 2021. Available from: https://www.nice.org.uk/guidance/ta679. Date accessed: 11 AUG 2021.
7. Packer M, Anker SD, Butler J, Filippatos G, Pocock SJ, Carson P, et al. Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure. *New England Journal of Medicine* 2020; **383**: 1413-24.
8. IQVIA. LPD, IQVIA Ltd, incorporating data derived from THIN, A Cegedim Database, May 2021.
9. NICE. Technology appraisal guidance [TA398]: Lumacaftor–ivacaftor for treating cystic fibrosis homozygous for the F508del mutation. 2016. Available from: https://www.nice.org.uk/guidance/ta398/chapter/committee-discussion#cost-effectiveness-2. Date accessed: 04/06/2021.
10. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. Process and methods [PMG9]. 2013. Available from: https://www.nice.org.uk/process/pmg9/chapter/the-reference-case. Date accessed: August 2020.
11. Ndefo UA, Anidiobi NO, Basheer E, Eaton AT. Empagliflozin (Jardiance): A Novel SGLT2 Inhibitor for the Treatment of Type-2 Diabetes. *P & T : a peer-reviewed journal for formulary management* 2015; **40**: 364-8.
12. European Medicines Agency. Jardiance. 2014. Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/jardiance. Date accessed: October 2020.
13. National Institute for Health and Care Excellence. Canagliflozin, dapagliflozin and empagliflozin as monotherapies for treating type 2 diabetes Technology appraisal guidance [TA390]. 2016. Available from: https://www.nice.org.uk/guidance/ta390. Date accessed: July 2020.
14. National Institute for Health and Care Excellence. Empagliflozin in combination therapy for treating type 2 diabetes Technology appraisal guidance [TA336]. 2015. Available from: https://www.nice.org.uk/guidance/ta336. Date accessed: July 2020.
15. ClinicalTrials.gov. EMPagliflozin outcomE tRial in Patients With chrOnic heaRt Failure With Reduced Ejection Fraction (EMPEROR-Reduced). 2020. Available from: https://clinicaltrials.gov/ct2/show/NCT03057977. Date accessed: August 2020.
16. McMurray JJV, Solomon SD, Inzucchi SE, Kober L, Kosiborod MN, Martinez FA, et al. Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction. *N Engl J Med* 2019; **381**: 1995-2008.

17. Boehringer Ingelheim International GmbH. A phase III randomised, double-blind trial to evaluate efficacy and safety of once daily empagliflozin 10 mg compared to placebo, in patients with chronic Heart Failure with reduced Ejection Fraction (HFrEF). Clinical Trial Report c28576542-012020. 2020.
18. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009; **150**: 604-12.
19. Rogers JK, Yaroshinsky A, Pocock SJ, Stokar D, Pogoda J. Analysis of recurrent events with an associated informative dropout time: Application of the joint frailty model. *Statistics in Medicine* 2016; **35**: 2195-205.
20. Zannad F, Ferreira JP, Pocock SJ, Zeller C, Anker SD, Butler J, et al. Cardiac and Kidney Benefits of Empagliflozin in Heart Failure Across the Spectrum of Kidney Function: Insights From EMPEROR-Reduced. *Circulation* 2021; **143**: 310-21.
21. Mayne TJ, Nordyke RJ, Schold JD, Weir MR, Mohan S. Defining a minimal clinically meaningful difference in 12-month estimated glomerular filtration rate for clinical trials in deceased donor kidney transplantation. *Clin Transplant* 2021; **35**: e14326.
22. Butler J, Anker SD, Filippatos G, Khan MS, Ferreira JP, Pocock SJ, et al. Empagliflozin and health-related quality of life outcomes in patients with heart failure with reduced ejection fraction: the EMPEROR-Reduced trial. *Eur Heart J* 2021.
23. Spertus J, Peterson E, Conard MW, Heidenreich PA, Krumholz HM, Jones P, et al. Monitoring clinical changes in patients with heart failure: a comparison of methods. *Am Heart J* 2005; **150**: 707-15.
24. Gov.uk. Population of England and Wales. 2020. Available from: https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/population-of-england-and-wales/latest. Date accessed: March.
25. Lam CSP, Ferreira JP, Pfarr E, Sim D, Tsutsui H, Anker SD, et al. Regional and ethnic influences on the response to empagliflozin in patients with heart failure and a reduced ejection fraction: the EMPEROR-Reduced trial. *Eur Heart J* 2021.
26. Zannad F, Ferreira JP, Pocock SJ, Anker SD, Butler J, Filippatos G, et al. SGLT2 inhibitors in patients with heart failure with reduced ejection fraction: a meta-analysis of the EMPEROR-Reduced and DAPA-HF trials. *Lancet* 2020; **396**: 819-29.
27. National Institute for Health and Care Excellence. Sacubitril valsartan for treating symptomatic chronic heart failure with reduced ejection fraction [TA388]. 2016. Available from: https://www.nice.org.uk/guidance/ta388. Date accessed.
28. Scottish Medicines Consortium (SMC). Sacubitril/valsartan (Entresto). 2016. Available from: https://www.scottishmedicines.org.uk/medicines-advice/sacubitrilvalsartan-entresto-fullsubmission-113216/. Date accessed.
29. Scottish Medicines Consortium (SMC). ivabradine (Procoralan). 2012. Available from: https://www.scottishmedicines.org.uk/medicines-advice/ivabradine-procoralan-fullsubmission-80512/. Date accessed.
30. Scottish Medicines Consortium (SMC). eplerenone (Inspra). 2012. Available from: https://www.scottishmedicines.org.uk/medicines-advice/eplerenone-inspra-fullsubmission-79312/. Date accessed.
31. All Wales Medicines Strategy Group (AWMSG). Eplerenone (Inspra) 25mg and 50mg film-coated tablets. Penarth: All Wales Therapeutics and Toxicology Centre (AWTTC), secretariat of the All Wales Medicines Strategy Group (AWMSG). AWMSG Secretariat Assessment Report Advice No. 2712. 2012. Available from: https://www.crd.york.ac.uk/crdweb/ShowRecord.asp?LinkFrom=OAI&ID=32013000329&LinkFrom=OAI&ID=32013000329. Date accessed.

32. McEwan P, Darlington O, McMurray JJV, Jhund PS, Docherty KF, Bohm M, et al. Cost-effectiveness of dapagliflozin as a treatment for heart failure with reduced ejection fraction: a multinational health-economic analysis of DAPA-HF. *Eur J Heart Fail* 2020; **22**: 2147-56.
33. McMurray JJV, Trueman D, Hancock E, Cowie MR, Briggs A, Taylor M, et al. Cost-effectiveness of sacubitril/valsartan in the treatment of heart failure with reduced ejection fraction. *Heart* 2018; **104**: 1006.
34. Lee D, Wilson K, Akehurst R, Cowie MR, Zannad F, Krum H, et al. Cost-effectiveness of eplerenone in patients with systolic heart failure and mild symptoms. *Heart* 2014; **100**: 1681-7.
35. Sullivan PW, Ghushchyan VH. EQ-5D Scores for Diabetes-Related Comorbidities. *Value Health* 2016; **19**: 1002-8.
36. Kerr M, Bray B, Medcalf J, O'Donoghue DJ, Matthews B. Estimating the financial cost of chronic kidney disease to the NHS in England. *Nephrol Dial Transplant* 2012; **27 Suppl 3**: iii73-80.
37. Kent S, Schlackow I, Lozano-Kühne J, Reith C, Emberson J, Haynes R, et al. What is the impact of chronic kidney disease stage and cardiovascular disease on the annual cost of hospital care in moderate-to-severe kidney disease? *BMC Nephrol* 2015; **16**: 65.
38. Alva ML, Gray A, Mihaylova B, Leal J, Holman RR. The impact of diabetes-related complications on healthcare costs: new results from the UKPDS (UKPDS 84). *Diabetic medicine : a journal of the British Diabetic Association* 2015; **32**: 459-66.
39. National life tables: UK, 1980-1982 to 2017-2019 2020 [Internet]. 2020. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables.
40. Jesky MD, Dutton M, Dasgupta I, Yadav P, Ng KP, Fenton A, et al. Health-Related Quality of Life Impacts Mortality but Not Progression to End-Stage Renal Disease in Pre-Dialysis Chronic Kidney Disease: A Prospective Observational Study. *PLoS ONE* 2016; **11**: e0165675.
41. Personal Social Services Research Unit. Unit costs of health and social care 2020. 2020. Available from: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2020/. Date accessed: March 2021.
42. National health service. National Schedule of NHS costs: 2018/2019. 2019. Available from: https://www.england.nhs.uk/national-cost-collection/. Date accessed: November 2020.
43. Eurostat. Harmonised Index of Consumer Prices (HICP): European Commission. 2020. Available from: https://ec.europa.eu/eurostat/web/hicp. Date accessed.
44. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value Health* 2010; **13**: 509-18.

9 Appendices

9.1 Quality assessment

Table 54. Quality assessment of EMPEROR-Reduced (Adapted from CS, Table 17)

Question on trial design	Trial Acronym/number	
	EMPEROR-Reduced (NCT03057977)	
	Company assessment	ERG agrees or disagrees
Was randomisation carried out appropriately?	Yes. Randomisation was performed by using a permuted block design with a computer pseudo-random number generator.	✓
Was the concealment of treatment allocation adequate?	Yes. An Interactive Response Technology System (voice response or web response) was used to determine treatment assignment.	✓
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes. Demographic and patient characteristics were well balanced between the two treatment groups at baseline, and randomisation was stratified by geographical region, diabetes status and eGFR at screening.	✓
Were the care providers, participants, and outcome assessors blind to treatment allocation?	Yes. This was a double-blind study. An Endpoint Adjudication Committee evaluated all reported and potential clinical events in a manner blinded to the treatment assignment.	✓
Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	No. Proportion of patients who discontinued study treatment was low and well balanced between the two treatment groups.	✓
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No. All outcomes specified in the study protocol were reported in the clinical study report.	✓
Did the analysis include an ITT analysis? If so, was this appropriate? Were appropriate methods used to account for missing data?	Yes. Efficacy analysis were performed in the randomised set.	✓

Abbreviations: CSR, Clinical Study Report; ERG, Evidence Review Group; ITT, intention to treat; N/A, not applicable; RCT, randomised controlled trial.

9.2 Baseline characteristics for EMPEROR-Reduced

Table 55. Demographic and baseline characteristics (mean) of randomised participants in EMPEROR-Reduced trial (Reproduced from CS, Table 15 and Lam *et al.*2021[25])

Baseline characteristic*	Empagliflozin 10 mg	Placebo	Region				
			Latin America	North America	Europe	Asia	Other
Number of subjects	1863	1867	1286	425	1353	493	173
Age (years), mean (SD)	67.2±10.8	66.5±11.2	64.4	68.6	69.6	66.3	60.4
Female sex, No (%)	437 (23.5)	456 (24.4)	398 (30.9)	98 (23.1)	262 (19.4)	101 (20.5)	34 (19.7)
Race, No (%)†							
White	1325 (71.1)	1304 (69.8)	1026 (79.8)	301 (70.8)	1281 (94.7)	0	21 (12.1)
Black	123 (6.6)	134 (7.2)	154 (12.0)	100 (23.5)	3 (0.2)	0	0
Asian	337 (18.1)	335 (17.9)	4 (0.3)	17 (4.0)	8 (0.6)	493 (100)	150 (86.7)
Other or missing	78 (4.2)	94 (5.0)	102 (7.9)	7 (1.6)	61 (4.5)	0	2 (1.2)
Region, No (%)							
North America	212 (11.4)	213 (11.4)					
Latin America	641 (34.4)	645 (34.5)					
Europe	676 (36.3)	677 (36.3)					
Asia	248 (13.3)	245 (13.1)					
Other	86 (4.6)	87 (4.7)					
NYHA functional class, No (%)							
II	1399 (75.1)	1401 (75.0)	NR	NR	NR	NR	NR

III	455 (24.4)	455 (24.4)	299 (23.3)	119 (28.0)	373 (27.6)	97 (19.7)	42 (24.3)
IV	9 (0.5)	11 (0.6)					
Body-mass index‡ (kg/m^2), mean (SD)	28.0±5.5	27.8±5.3	28.1	29.6	28.9	24.0	24.6
Heart rate (beats/min), mean (SD)	71.0±11.7	71.5±11.8	71.0	70.4	70.9	72.6	74.9
SBP (mm Hg), mean (SD)	122.6±15.9	121.4+15.4	121.1	118.8	124.9	119.3	121.6
DBP (mm Hg), mean (SD)	74.0 (11.0)	73.7 (10.6)	NR	NR	NR	NR	NR
Left ventricular ejection fraction							
Mean (SD)	27.7±6.0	27.2±6.1	27.5	26.3	27.5	28.4	27.2
Value of ≤ 30%, No (%)	1337 (71.8)	1392 (74.6)	NR	NR	NR	NR	NR
NT pro-BNP							
Median (IQR) (pg/ml)	1887 (1077-3429)	1926 (1153-3525)	3286 (1096, 3819)	2868 (1072, 3297)	2820 (1147, 3212)	3153 (1271, 3693)	2913 (999, 3076)
Value of ≥1000 pg/ml, No/total No (%)	1463/1862 (78.6)	1488/1866 (79.7)	NR	NR	NR	NR	NR
Cause of heart failure, No (%)							
Ischaemic	983 (52.8)	946 (50.7)	NR	NR	NR	NR	NR
Nonischaemic	880 (47.2)	921 (49.3)	NR	NR	NR	NR	NR
Cardiovascular history, No (%)							

Hospitalisation for HF in ≤12 months	577 (31.0)	574 (30.7)	314 (24.4)	134 (31.5)	417 (30.8)	238 (48.3)	48 (27.7)
Atrial fibrillation	664 (35.6)	705 (37.8)	328 (25.5)	187 (44.0)	658 (48.6)	170 (36.3)	17 (9.8)
Diabetes mellitus	927 (49.8)	929 (49.8)	625 (48.6)	222 (52.2)	660 (48.8)	244 (49.5)	105 (60.7)
Hypertension	1349 (72.4)	1349 (72.3)	922 (71.7)	345 (81.2)	1022 (75.5)	319 (64.7)	90 (52.0)
Estimated glomerular filtration rate							
Mean (SD) (ml/min/1.73 m^2)	61.8 ± 21.7	62.2 ± 21.5	64.7	58.7	57.9	65.6	71.8
Value of <60 ml/min/1.73 m^2, No/total No (%)	893/1862 (48.0)	906/1866 (48.6)	NR	NR	NR	NR	NR
UACR (mg/ml), N (%)							
Normal (<30)	1038 (55.7)	1040 (55.7)	NR	NR	NR	NR	NR
Microalbuminuria (30 to ≤300)	608 (32.6)	628 (33.6)	NR	NR	NR	NR	NR
Macroalbuminuria (>300)	207 (11.1)	189 (10.1)	NR	NR	NR	NR	NR
Heart failure medication, No (%)							
Renin–angiotensin inhibitor§							
Without neprilysin Inhibitor	1314 (70.5)	1286 (68.9)	1008 (78.4)	206 (48.5)	971 (71.8)	330 (66.9)	85 (49.1)
With neprilysin Inhibitor	340 (18.3)	387 (20.7)	178 (13.8)	161 (37.9)	276 (20.4)	67 (13.6)	45 (26.0)
Mineralocorticoid receptor antagonist	1306 (70.1)	1355 (72.6)	1032 (80.2)	210 (49.4)	975 (72.1)	324 (65.7)	120 (69.4)

Beta-blocker	1765 (94.7)	1768 (94.7)	1235 (96.0)	406 (95.5)	1302 (96.2)	452 (91.7)	138 (79.8)
Device therapy, No (%)							
Implantable cardioverter-defibrillator¶	578 (31.0)	593 (31.8)	94 (7.3)	273 (64.2)	718 (53.1)	67 (13.6)	18 (10.4)
Cardiac resynchronisation therapyǁ	220 (11.8)	222 (11.9)	39 (3.0)	84 (19.8)	272 (20.1)	38 (7.7)	5 (2.9)
Diabetes status							
Without diabetes, N (%)	936 (50.2)	938 (50.2)	NR	NR	NR	NR	NR
Without diabetes or pre-diabetes, N (%)	304 (16.3)	302 (16.2)	NR	NR	NR	NR	NR
With pre-diabetes, N (%)	632 (33.9)	636 (34.1)	NR	NR	NR	NR	NR
With diabetes, N (%)	927 (49.8)	929 (49.8)	625 (48.6)	222 (52.2)	660 (48.8)	244 (49.5)	105 (60.7)
T2DM, N (%)	927 (49.8)	929 (49.8)	NR	NR	NR	NR	NR
T1DM, N (%)	0	0	NR	NR	NR	NR	NR

Abbreviations: DBP, diastolic blood pressure; HF, heart failure; IQR, interquartile range; No, number; NR, not reported; NT pro-BNP, N-terminal prohormone of brain natriuretic peptide; NYHA, New York Heart Association; SBP, systolic blood pressure; SD, standard deviation; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus; UACR, urine albumin-to-creatinine ratio.

* Plus-minus values are means ± SD. Percentages may not total 100 because of rounding.

† Race was reported by the patients. Those who identified with more than one race or with no race were classified as “other”.

‡ The body-mass index is the weight in kilograms divided by the square of the height in meters.

§ Inhibitors of the renin–angiotensin system include angiotensin converting enzyme inhibitors and angiotensin receptor blockers.

¶ This category includes all the patients with an implantable cardioverter-defibrillator regardless of the presence or absence of cardiac resynchronisation therapy.

ǁ This category includes all the patients who were receiving cardiac resynchronisation therapy regardless of the presence or absence of a defibrillator.

9.3 Summary of the study's included in the ITC

Table 56.Summary of the trials used to carry out the indirect treatment comparison (Reproduced from CS, Table 22)

Trial reference	EMPEROR-Reduced	DAPA-HF
Intervention (N)	Empagliflozin (10 mg qd) + SoC	Dapagliflozin (10 mg or 5 mg qd) + SoC
Comparator (N)	Placebo + SoC	Placebo + SoC
Study start completion (years)	2017–2020	2017–2019
Phase	III	III
Method of blinding	Double-blind	Double-blind
Randomisation	1:1, stratified by geographical region, history of diabetes and eGFR	1:1, stratified by type II diabetes (with and without)
Study centres	Multicentre (Europe, North America, Latin America, Asia, Other)	Multicentre (Europe, North America, Latin America, Asia Pacific)
Primary composite	The composite primary endpoint for this trial was the time to first event of adjudicated CV death or adjudicated HHF	Time to the first occurrence of any of either CV death, hospitalisation for HF or an urgent HF visit
Secondary outcomes	Key secondary outcomes: • Occurrence of adjudicated HHF (first and recurrent) • eGFR $(CKD\text{-}EPI)_{cr}$ slope of change from baseline Other secondary outcomes: • Time to the first event in the composite renal endpoint: chronic dialysis, renal transplant, or sustained reduction in eGFR $(CKD\text{-}EPI)_{cr}$ • Time to first adjudicated HHF • Time to adjudicated CV death • Time to all-cause mortality • Time to onset of T2DM in patients with pre-T2DM • Change from baseline in KCCQ clinical summary at week 52 • Occurrence of all-cause hospitalisation (first and recurrent)	• Time to the first occurrence of CV death or hospitalisation for HF • Total number of (first and recurrent) HF hospitalisations and CV death • Change from baseline measured at 8 months in KCCQ overall summary score • Renal composite: ≥50% sustained decline in eGFR, reaching end-stage renal disease or renal death • Time to death from any cause
Median follow-up duration	16 months	18.2 months

Abbreviations: N, number of participants; qd, once a day; SoC, standard of care.

National Institute for Health and Care Excellence

Centre for Health Technology Evaluation

ERG report – factual accuracy check and confidential information check

Empagliflozin for treating chronic heart failure with reduced ejection fraction [ID3826]

'*Data owners will be asked to check that confidential information is correctly marked in documents created by others in the technology appraisal process before release; for example, the technical report and ERG report.*' (Section 3.1.29, Guide to the processes of technology appraisals).

You are asked to check the ERG report to ensure there are no factual inaccuracies or errors in the marking of confidential information contained within it. The document should act as a method of detailing any inaccuracies found and how they should be corrected.

If you do identify any factual inaccuracies or errors in the marking of confidential information, you must inform NICE by **5pm on Tuesday 14 September 2021** using the below comments table.

All factual errors will be highlighted in a report and presented to the Appraisal Committee and will subsequently be published on the NICE website with the committee papers.

Please underline all confidential information, and separately highlight information that is submitted as 'commercial in confidence' in turquoise, all information submitted as 'academic in confidence' in yellow, and all information submitted as 'depersonalised data' in pink.

Issue 1 Source data for the prescribing of dapagliflozin in heart failure

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 34 of the ERG report states "the dapagliflozin recommendation requires the involvement of a HF specialist to initiate treatment and that it was not available at the time of publication of NG106. The company argue that dapagliflozin should not be included as a comparator for empagliflozin as it's uptake also remains low, although the ERG and the ERG's clinical experts disagree and consider dapagliflozin use to be increasing and that it is a relevant comparator.	Data from the THIN database showed that prescribing of dapagliflozin in heart failure is low and therefore it is not a relevant comparator (Document B, Table 2). The statement that clinical experts disagree and consider dapagliflozin use to be increasing and is a relevant comparator is not substantiated with any data.	Substantiation of a statement	Not a factual inaccuracy. No change required.

Issue 2 Rationale for cost comparison case

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 40 of the ERG report states "The company's model adheres to the decision problem for the comparison of empagliflozin and SoC. However, the company conducted a cost-comparison analysis for empagliflozin vs dapagliflozin, which the ERG disagrees with. The ERG recommends that the company uses the results from the Bucher	Suggested amendment: The company's model adheres to the decision problem for the comparison of empagliflozin and SoC. However, the company conducted a cost-comparison analysis for empagliflozin vs dapagliflozin, which the ERG disagrees with. The ERG recommends that the company uses the results from the Bucher analysis to conduct a cost utility analysis for these drugs.	The text provides additional clarification on why the Company has presented a cost comparison case.	Not a factual inaccuracy. No change required.

analysis to conduct a cost utility analysis for these drugs."	*The company noted in the CS that the conclusion from the Bucher ITC that empagliflozin and dapagliflozin offer comparable efficacy across key outcomes for patients with HFrEF is consistent with feedback from UK clinical experts (CS, Page 103).* *The company maintains that the criteria for a cost comparison, as detailed in the Guide to the Processes of Technology Appraisal 2013, are met.*		

Issue 3 Provision of outcome data for the Europe subgroup

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 19 of the ERG report states that "baseline characteristics are required for the Europe subgroup to enable a comparison with the ITT population of EMPEROR-R". "In addition, results for all-cause mortality, KCCQ-CSS and renal function for the Europe subgroup should be presented."	Suggested amendment: *During clarification questions, the manufacturer referred to a publication on the outcomes of EMPEROR-R by region (Lam et al 2021). This publication reports on the baseline characteristics and outcomes in the Europe subgroup and discusses why there might be differences between regions.* *The manufacturer noted that the NICE committee for dapagliflozin (TA679) expressed concerns that the Europe subgroup in DAPA-HF did not reflect the diversity or the UK population, as patients were predominantly white. The same trend was observed for EMPEROR-Reduced. Thus, if the Europe subgroup*	• Data on the expected outcomes for the Europe subgroup has already been provided by the Company to support the ERGs assessment of the submission • Description of why the manufacturer did not present this information in the company submission.	The ERG thanks the company for highlighting the availability of the baseline characteristics for the Europe subgroup in the Lam *et al.* 2021 publication and has added these to the ERG report. The text in the ERG report has also been updated to incorporate the ERG's view on the Europe baseline characteristics.

	was used as the base case for the committee to base their recommendation, there is a risk it is inconsistent with NICE's Social Value Judgments and the Equality Act 2010. Reference: Lam CSP, Ferreira JP, Pfarr E, Sim D, Tsutsui H, Anker SD, et al. Regional and ethnic influences on the response to empagliflozin in patients with heart failure and a reduced ejection fraction: the EMPEROR-Reduced trial. Eur Heart J. 2021.[https://doi.org/10.1093/eurheartj/ehab360]		

Issue 4 Number of CV-death events

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 23 of the ERG report states "the company added that CV deaths were rare events in EMPEROR-R".	Suggested amendment: "The company added that CV deaths were relatively rare *with around 10% of patients dying due to a CV event. With low event counts, the observed shape of the KM curves must be interpreted carefully to distinguish between chance variation and true signals of change*".	The suggested amendment provides additional information on what is meant by rare and to allow the reader to make an objective inference.	The ERG has partially amended its statement to, "*The company added that CV deaths were relatively rare with around 10% of patients dying due to a CV event"* to reflect the company's view more accurately in the text.

Issue 5 Number of predicted CV-death events in the CE model

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 24 of the ERG report states "in PULSE, there were 7,905 CV deaths and 9,599 non-CV deaths over a mean follow-up of 3 years. In the model, there were [redacted] CV deaths and [redacted] non-CV deaths when the subgroup data from EMPEROR-R is used."	Please state whether the predicted number of CV-death and non-CV death events [redacted] and [redacted]) in the CE model was also over a 3-year time horizon, similar to PULSE.	Ensures a like for like comparison on the predicted vs observed number of death events in the CE model vs PULSE, respectively.	As reported on page 119 of the ERG report, the estimates obtained by the ERG are for a period of 3 years in the model. The ERG has added text on page 24, as requested by the company.

Issue 6 Company rationale for presenting a Bucher ITC for empagliflozin vs dapagliflozin

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 36 of the ERG report states "Lastly, as discussed above (Section **Error! Reference source not found.**), the ERG and its clinical experts disagree with the company's proposal that dapagliflozin is not a relevant comparator. However, the ERG notes that the company has provided results from an ITC of empagliflozin vs dapagliflozin in the CS."	Suggested amendment: "Lastly, as discussed above (Section **Error! Reference source not found.**), the ERG and its clinical experts disagree with the company's proposal that dapagliflozin is not a relevant comparator." *The Company noted in the CS that the ITC of empagliflozin vs dapagliflozin was provided upon the request of NICE and the first ERG at the Decision Problem Meeting."*	Text accurately reflects prior discussions at the Decision Problem Meeting for this appraisal	Not a factual inaccuracy. No change required

Issue 7 Use of KCCQ to capture disease progression in the CE model

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 40 of the ERG report states "The ERG also notes that the KCCQ was used in EMPEROR-Reduced to capture health status and that one of its domains includes symptoms. The ERG's clinical experts reported the KCCQ is a reasonable tool for assessing symptoms, although it doesn't tend to be routinely used in clinical practice."	Suggested amendment: *In the CS, the company presented the rationale for using KCCQ. This included the questionnaire's robustness and ability to quantify the patient's perception of their health status including HF symptoms (unlike the NYHA classification which is a physician's interpretation of patient's symptoms). Furthermore, the KCCQ was used in DAPA-HF study and in the dapagliflozin appraisal (TA679). In the submission for TA679, it is noted that 'KCCQ rather than NYHA class, has become the standard tool used in clinical trials to evaluate patient-reported health status and response to treatment.*"	Provides a balance of perspectives. The text acknowledges that while the ERG's clinical experts noted the KCCQ doesn't tend to be routinely used in clinical practice, there is a clear rationale for its use, which was noted in the CS, and it was accepted in another technology appraisal for the same drug class.	Not a factual inaccuracy. No change required

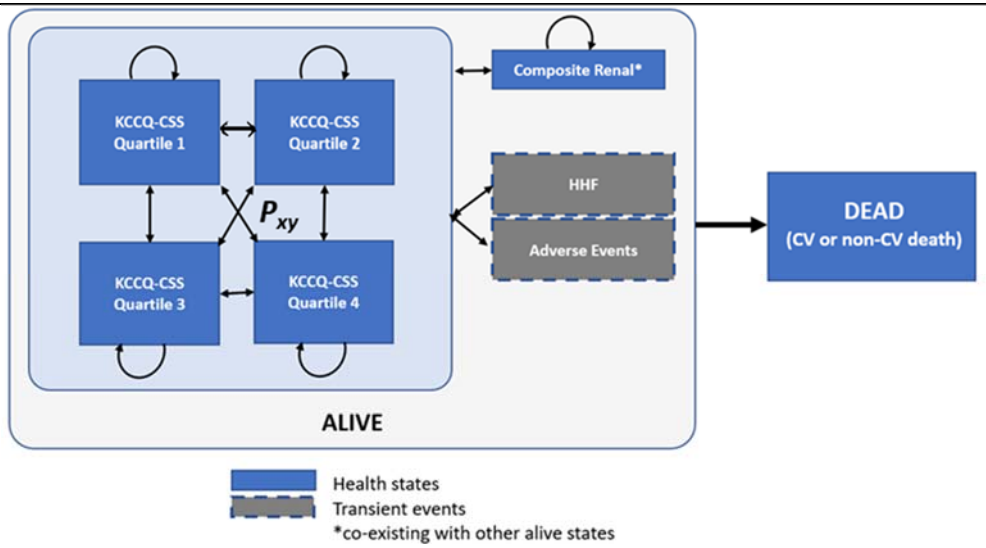
Issue 8 Updated regulatory information

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 45 of the ERG report states "The company also reported that a submission had been made to the Medicines and Healthcare products Regulatory Agency	Suggested amendment: *"On 30 July 2021, the Medicines and Healthcare products Regulatory Agency (MHRA) also approved empagliflozin for the indication of relevance to this technology appraisal. The indication wording is "Jardiance is indicated in adults for the*	Provision of updated regulatory information	The ERG thanks the company for the update and has updated the ERG report to reflect the updated information

(MHRA), via the reliance route, on z xxxxx xxxx and UK MHRA Marketing Authorisation for empagliflozin use in HFrEF is expected in the week commencing [redacted]. The ERG was unable to find an update on the MHRA website as of [redacted]."	*treatment of symptomatic chronic heart failure with reduced ejection fraction. The Jardiance Summary of Product Characteristics (SmPC) was revised on 30 July 2021 and includes the heart failure indication. The SmPC is available on the MHRA website (https://products.mhra.gov.uk/search/?search=jardiance&page=1) and also on www.medicines.org.uk"*		provided by the company.

Issue 9 Updated model diagram

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Figure 13 (Page 95) of the ERG report contains a minor error. An arrow is missing between KCCQ-CSS Quartile 1 and KCCQ-CSS Quartile 2.	Include a revised figure with a double (left-right) arrow included between KCCQ-CSS Quartile 1 and KCCQ-CSS Quartile 2, to ensure it is clear that patients may transition between these two quartiles.	To ensure the model structure is accurately represented.	The ERG thanks the company for providing a corrected model schematic. This has been updated in the ERG report.

Issue 10 Update Table number and source document

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The company reported that Table 22, "shows that the proportion of patients who have more than 5 or 10 decline in eGFR from baseline is comparable for the empagliflozin and placebo arm".	Suggested amendment: The company reported that *Table 4* of the *clarification questions*, "shows that the proportion of patients who have more than 5 or 10 decline in eGFR from baseline is comparable for the empagliflozin and placebo arm".	Change the Table number and include a reference to the source document	The ERG has amended the ERG report to accurately reflect the source document.

Issue 11 KCCQ-CSS data (randomised set) has already been provided by the Company

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The ERG's preferred data set is the randomised set for both the baseline and 52 week results, although this was not provided and so the ERG discusses the data as provided by the company (**Error! Reference source not found.**).	Suggested amendment: The baseline KCCQ-CSS scores for the randomised set were provided in Table 8 of the Company Clarification Questions response and are reported in the Clinical Trial Report (Tables 15.2.3.6:5 and 15.2.3.6:1).	Data has already been provided by the Company to support the ERGs assessment of the submission	The ERG thanks the company for highlighting this error and has amended the text and content of Table 24 to discuss the results from the randomised set.

Issue 12 Addition of confidence intervals to HHF data

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
In particular, the ERG notes that the benefit with empagliflozin compared to placebo in reducing total HHF events is only statistically significant in [redacted]	Suggested adding confidence intervals to point estimates "*In particular, the ERG notes that the benefit with empagliflozin compared to placebo in* [redacted]	BI suggests adding confidence intervals to allow the reader to form their own interpretation about HHF outcomes by the age subgroup.	Not a factual inaccuracy. No change required.

Issue 13 Clarity on modelling approach for HHF

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Given that time to HHF Kaplan-Meier (KM) data were available from EMPEROR-R, the ERG considers that the company could have used these data to model time to HHF. Furthermore, using KM data for time to HHF would have allowed the company to model time to first and subsequent HHF separately. Therefore, the ERG considers that the company's approach to modelling HHF could have relied on more appropriate methods.	Kaplan Meir data is only for time to first event and does capture subsequent HHF events.	Additional clarity on the proposed modelling approach	Not a factual inaccuracy. No change required. Table 15.2.4.2:1 in the CSR presents KM data from first HHF to second HHF.

Issue 14 Spelling error

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The lack of flexibility in the company's model structure and the lack of **KCQQ-**CSS data from PULSE mean that the company cannot use mortality data from	Amend to: The lack of flexibility in the company's model structure and the lack of ***KCCQ-CSS*** data from PULSE mean that the company cannot use	Spelling error (KCQQ to KCCQ).	The ERG thanks the company for highlighting this error. This has been corrected in the ERG report.

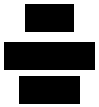
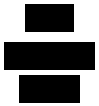
PULSE directly in the SoC arm of the model	mortality data from PULSE directly in the SoC arm of the model		

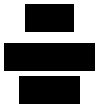
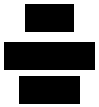
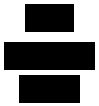
Issue 15 Data for the duration of HHF has already been provided by the Company

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 124 of the ERG report states "The ERG notes that it did not have access to mean (or median) duration of hospitalisation in EMPEROR-R, and so it cannot ascertain the extent to the overestimation of this disutility."	Duration of hospitalisation is provided in Table 15.2.4.1:1 and 15.2.4.33: 2 of the Clinical Trial Report	Data has been provided by the Company to support the ERGs assessment of the submission	The ERG thanks the company for the additional information. The ERG asks that the company clarifies the time units for the estimates provided in Table 15.2.4.33:2, where it is reported that the mean HHF stay in the trial was xxxx (i.e., days, weeks, months, etc).

Location of incorrect marking	Description of incorrect marking	Amended marking	ERG response
ID3826 Empagliflozin for heart failure ERG report 020921 GK [ACIC], page 19, table 4	Details of the Bucher ITC should be CIC and not AIC. This is commercially sensitive and is likely of interest to competitors who may be commentators on the Technical Engagement.	The ERG is concerned that the results of the Bucher ITC show a trend suggesting [redacted] CV deaths and all-cause mortality with empagliflozin compared to dapagliflozin (HR [empagliflozin vs dapagliflozin] [redacted], respectively) and a [redacted] mean change in KCCQ-CSS score from baseline at 8 months with dapagliflozin (MD xxx) compared with empagliflozin (MD [redacted]). The results from the ITC are not used in the economic model and instead the company has assumed a class effect for SGLT2is. The ERG considers the results of the Bucher ITC to be [redacted] and that the company's assumption of equal effectiveness for empagliflozin and dapagliflozin lacks robustness.	The ERG thanks the company for highlighting this error. The confidential marking has been corrected in the ERG report.
ID3826 Empagliflozin for heart failure ERG report 020921 GK [ACIC], page 3, table 18, (comparators row)	Details of the Bucher ITC should be CIC and not AIC. This is commercially sensitive and is likely of interest to competitors who may be commentators on the Technical Engagement.	The ERG is concerned that the company is making a strong assumption of equivalence for the clinical-effectiveness of empagliflozin and dapagliflozin based on a single trial for each drug, with [redacted] results from the ITC. The ERG thus considers the results of the pooled meta-analysis conducted by the company, where it is assuming a class effect for SGLT2is, should be interpreted with caution and instead prefers the use of the Bucher ITC.	The ERG thanks the company for highlighting this error. The confidential marking has been corrected in the ERG report.
ID3826 Empagliflozin for heart failure ERG report 020921 GK [ACIC], page 77,	Details of the Bucher ITC should be CIC and not AIC. This is commercially sensitive and is likely of interest to	The results of the Bucher ITCs showed [redacted] between empagliflozin and dapagliflozin (**Error! Reference source not found.**). However, there was a trend suggesting [redacted] CV deaths and all-cause mortality with empagliflozin compared to	The ERG thanks the company for highlighting this error. The confidential marking has been corrected in the ERG report.

<table>
<tr>
<td>section 3.4.2 ‘Results of the Bucher ITCs’</td>
<td>competitors who may be commentators on the Technical Engagement.</td>
<td>dapagliflozin (HR [empagliflozin vs dapagliflozin] [redacted] respectively). In addition, there appears to be a [redacted] mean change in KCCQ-CSS score from baseline at 8 months with dapagliflozin (mean difference [MD] xxx) compared with empagliflozin (MD xxx), albeit potentially [redacted] The ERG notes that neither study was powered to detect differences in these specific outcomes as they were both powered for their own trial specific primary composite outcomes.</td>
<td></td>
</tr>
<tr>
<td>ID3826 Empagliflozin for heart failure ERG report 020921 GK [ACIC] page 78</td>
<td>Details of the Bucher ITC should be CIC and not AIC. This is commercially sensitive and is likely of interest to competitors who may be commentators on the Technical Engagement.</td>
<td>The Bucher ITC of worsening renal function using the DAPA-HF definition suggests that empagliflozin might be [redacted] in reducing the hazard of worsening renal function compared to dapagliflozin, although the [redacted] (HR [redacted]; 95% CI: [redacted]).</td>
<td>The ERG thanks the company for highlighting this error. The confidential marking has been corrected in the ERG report.</td>
</tr>
<tr>
<td>ID3826 Empagliflozin for heart failure ERG report 020921 GK [ACIC], page 78, table 31</td>
<td>Details of the Bucher ITC should be CIC and not AIC. This is commercially sensitive and is likely of interest to competitors who may be commentators on the Technical Engagement.</td>
<td>
<table>
<tr><th>Endpoint: relative effect measure</th><th>EMPEROR-REDUCED: empagliflozin versus placebo[a]</th><th>DAPA-HF: dapagliflozin versus placebo[a]</th><th>Bucher ITC: empagliflozin versus dapagliflozin[a]</th></tr>
<tr><td>Time to first event of adjudicated CV death or adjudicated HHF: HR (95% CI)</td><td>0.75
(0.65 to 0.86)</td><td>0.75
(0.65 to 0.85)</td><td>[redacted]</td></tr>
</table>
</td>
<td>The ERG thanks the company for highlighting this error. The confidential marking has been corrected in the ERG report.</td>
</tr>
</table>

		Time to first event of adjudicated CV death or adjudicated HHF (EMPEROR-Reduced) versus Time to first worsening of heart failure (hospitalisation or an urgent visit resulting in intravenous therapy for heart failure) or CV death (DAPA-HF): HR (95% CI)	0.75 (0.65 to 0.86)	0.74 (0.65 to 0.85)		
		Time to first adjudicated HHF: HR (95% CI)	0.69 (0.59 to 0.81)	0.70 (0.59 to 0.83)		
		Time to adjudicated CV death: HR (95% CI)	0.92 (0.75 to 1.12)	0.82 (0.69 to 0.98)		
		Time to all-cause	0.92 (0.77 to 1.1)	0.83 (0.71 to 0.97)		

		mortality: HR (95% CI)				
		Occurrence of adjudicated HHF (first and recurrent) – analysed using a joint frailty model: HR (95% CI)	0.70 (0.58 to 0.85)	0.71 (0.61 to 0.82)		
		Occurrence of adjudicated HHF (first and recurrent) – analysed using a Lin-Wei-Yang-Ying model: RR (95% CI)	0.76 (0.65 to 0.89)	0.75 (0.65 to 0.88)		
		Worsening renal function (as defined in DAPA-HF): HR (95% CI)	0.52 (0.29 to 0.92)	0.71 (0.44 to 1.16)		
		Change in KCCQ total symptom score at 8 months/7.4 months: MD (SE/95% CI)	1.6 (0.7)	2.8 (0.5)		

		Abbreviations: CV, cardiovascular; HHF, hospitalization for heart failure; HR hazard ration; CI, confidence interval; RR, risk ratio; KCCQ, Kansas City Cardiomyopathy Questionnaire.	
ID3826 Empagliflozin for heart failure ERG report 020921 GK [ACIC], page 86	Details of the Bucher ITC should be CIC and not AIC. This is commercially sensitive and is likely of interest to competitors who may be commentators on the Technical Engagement.	The results of the Bucher ITC showed [redacted] between empagliflozin and dapagliflozin. However, there was a trend suggesting xxxxxx CV deaths and all-cause mortality with empagliflozin compared to dapagliflozin (HR [empagliflozin vs dapagliflozin] [redacted], respectively) and a [redacted] mean change in KCCQ-CSS score from baseline at 8 months with dapagliflozin (mean difference [MD] xxx) compared with empagliflozin (MD xxx). The ERG notes that the results of the ITC are not used in the economic model, and the company instead assumes equal clinical-effectiveness for empagliflozin and dapagliflozin. The ERG is concerned that the company is making a strong assumption of equivalence for empagliflozin and dapagliflozin based on a single trial for each drug, with [redacted] results from the ITC. The ERG thus considers the results of the pooled meta-analysis conducted by the company, where it is assuming a class effect for SGLT2is, should be interpreted with caution.	The ERG thanks the company for highlighting this error. The confidential marking has been corrected in the ERG report.
ID3826 Empagliflozin for heart failure ERG report 020921 GK [ACIC], page 94		As discussed in Section 3.4 and 3.5, the ERG is concerned that the company is making a strong assumption of equivalence for empagliflozin and dapagliflozin based on a single trial for each drug, with [redacted] results from the ITC and therefore, considers the company's Bucher ITC a more appropriate method of assessing the efficacy of empagliflozin vs dapagliflozin for this technology appraisal.	The ERG thanks the company for highlighting this error. The confidential marking has been corrected in the ERG report.

NICE National Institute for Health and Care Excellence

Technical engagement response form

Empagliflozin for treating chronic heart failure with reduced ejection fraction [ID3826]

As a stakeholder you have been invited to comment on the ERG report for this appraisal. The ERG report and stakeholders' responses are used by the appraisal Committee to help it make decisions at the appraisal Committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the key issues below. You do not have to provide a response to every issue. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be included in the Committee papers in full and may also be summarised and presented in slides at the appraisal Committee meeting.

Deadline for comments by **5pm** on **21 October 2021**

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the ERG report which summarises the background and submitted evidence, and presents the ERG's summary of key issues, critique of the evidence and exploratory analyses. This will provide context and describe the questions below in greater detail.
- Please ensure your response clearly identifies the issue numbers that have been used in the executive summary of the ERG report. If you would like to comment on issues in the ERG report that have not been identified as key issues, you can do so in the 'Additional issues' section.
- If you are the Company involved in this appraisal, please complete the 'Summary of changes to the Company's cost-effectiveness estimates(s)' section if your response includes changes to your cost-effectiveness evidence.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response unreadable. Please type information directly into the form.
- Do not include medical information about yourself or another person that could identify you or the other person.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.
- If you provide journal articles to support your comments, you must have copyright clearance for these articles.

- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow, and all information submitted under 'depersonalised data' in pink. If confidential information is submitted, please also send a second version of your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received during engagement are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

NICE National Institute for Health and Care Excellence

About you

Your name	[redacted]
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	**Boehringer Ingelheim**
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	**None**

Key issues for engagement

Please use the table below to respond to questions raised in the ERG report on key issues. You may also provide additional comments on the key issue that you would like to raise but which do not address the specific questions.

Key Issue	Does this response contain new evidence, data or analyses?	Response
Key Issue 1: Uncertainty around the generalisability of the results from EMPEROR-Reduced to the older heart failure with reduced ejection fraction population expected in clinical practice	YES	• **The baseline characteristics in the ITT and >65 years subgroup in EMPEROR-R were broadly comparable.** • **There is limited evidence that the >65 years subgroup is a better representation of a typical UK patient. Use of the >65 years group would decrease certainty in the estimate of cost-effectiveness of empagliflozin vs SoC due to a lower sample size.** • **A conclusion about the generalisability of the >65 years subgroup EMPEROR-R to UK clinical practice is further hindered by their being a significant amount of missing data in some baseline characteristics in PULSE (e.g. BMI, ejection fraction, NT-proBNP, heart rate, eGFR), making it impossible to characterise a 'typical' UK patient.** • **Use of the >65 years subgroup would also increase disparity in access across ethnic and socio-economic groups. This is because HF patients in lower socio-economic groups and those belonging to ethnic minorities tend to be younger. Thus, the ITT population is the most suitable population for decision making.**

		The ERG expressed concerns about the generalisability of EMPEROR-R to the older HFrEF population expected in UK clinical practice. This is because the mean age in EMPEROR-R was 66.8 years (ITT population) compared to [redacted] years in the PULSE study. During technical engagement, the ERG sought further clarification on how representative the >65 years subgroup in EMPEROR-R was to UK clinical practice. There is limited evidence that the >65 years subgroup is a better representation of the typical UK patient than the ITT population. Most baseline characteristics of the >65 years subgroup were broadly comparable to the ITT population in EMPEROR-R (Table 1). These included the percentage with T2DM, the percentage with HHF in the 12 months prior to randomisation, the mean time since diagnosis, mean LVEF, mean and median NT-proBNP, mean blood pressure and mean BMI (kg/m^2). However, some differences were observed including the percentage with atrial fibrillation flutter and the percentage of patients with an eGFR ≥60ml/min.1.73m2 at baseline. As well as the >65 years subgroup having limited value over the ITT population in demonstrating generalisability to the UK HF population, it represents only 62% of the trial data. Therefore, its use as the preferred subgroup for the base case will only introduce greater uncertainty in the estimation of CV-mortality and hospitalisation for heart failure (HHF) (Issue 6 and 8), and subsequently the ICER. As stated in Appendix O and the Company Clarification Question 10 response, the outcomes in the ITT population and the >65 years group were comparable. This is evident in the cost per QALY still remaining well below the £20,000/QALY threshold (£7,717/QALY vs £7,213/QALY, respectively).

		A conclusion about the generalisability of EMPEROR-R to UK clinical practice is further hindered by their being a significant amount of missing data in some baseline characteristics in PULSE (e.g. BMI, ejection fraction, NT-proBNP, heart rate, eGFR) (Table 1). Some of these factors, such as NT-proBNP and eGFR are prognostic and help to identify patients who are at 'higher risk' for worse cardiovascular outcomes.(1) Missing data is an inherent limitation of any real-world evidence study, commonly encountered when using the CPRD as a data source. Therefore, it is impossible to define what is a typical HF patient in the UK and draw a definite conclusion about the generalisability of the EMPEROR-R trial population. Preferring the >65 years subgroup represents an equity in access issue. If this subgroup is used as a basis to make a recommendation by the Committee, it could further exacerbate pre-existing equalities, as younger HFrEF patients tend to belong to lower socio-economic groups and ethnic minorities. As stated in NICE's guiding principles, any recommendation must be consistent with the Equality Act 2010(2) which has nine protected characteristics, including age and ethnicity. In a CRPD study of 4 million individuals in the UK, Conrad et al 2018 reported that socio-economically deprived individuals were more likely to develop HF than affluent individuals (incidence rate-ratio 1.61,95% CI 1.58 -1.64) and were 3.5 years younger at diagnosis (95% CI between 3.77 to 3.25 younger) (3). In PULSE, there was a high prevalence and incidence of heart failure in the North West ([redacted] an area of the UK with pockets of high deprivation. Younger patients were also more likely to belong to ethnic minorities. Lawson et al 2020 reported that south asians and the black group were six and nine years younger than whites at diagnosis of HF (4). In conclusion, adopting the >65 years subgroup as the preferred base case would ignore benefits that empagliflozin could offer for HF patients living in socio-

economically deprived areas of the UK and ethnic minorities, and would be inconsistent with the UK Government's levelling up agenda (5).

Since the outcomes in the >65 years subgroup were comparable to the ITT population, even if the >65 years subgroup was preferred by the committee, empagliflozin would be cost effective (£6,342/QALY vs £4,717/QALY in the original company base case).

Table 1. Baseline characteristics of EMPEROR-R (ITT, >65 years) vs PULSE

Treatment Arm	EMPEROR-R (ITT; Combined Groups)	EMPEROR-R (65+; Combined Groups)	PULSE[a]
N	3,730	2315	[redacted]
Age [mean (SD)]	66.8 (11.0)	73.8 (5.9)	[redacted]
Age at baseline >= 65 years [% (N)]	62.1% (2315)	100.0% (2315)	NR
Gender [% (N)]			
Male	76.1% (2837)	76.1% (1762)	[redacted]
Race [% (N)]			[redacted]
White	70.5% (2629)	75.6% (1750)	[redacted]
Asian	18.0% (672)	15.9% (369)	[redacted]
Native	1.0% (39)	0.9% (21)	[redacted]
Black	6.9% (257)	4.6% (106)	[redacted]
Multiple/mixed	1.6% (61)	1.0% (24)	[redacted]
Pacific	0.4% (14)	0.3% (8)	NR
Region [% (N)], global			
Asian	13.2% (493)	13.0% (301)	NR
Europe	36.3% (1353)	42.7% (988)	[redacted]

		Latin America	34.5% (1286)	29.2% (677)	NR
		North America	11.4% (425)	12.2% (282)	NR
		Other	4.6% (173)	2.9% (67)	NR
		Region [% (N)], UK			
		East Midlands	NR	NR	[redacted]
		East of England	NR	NR	[redacted]
		London	NR	NR	[redacted]
		North East	NR	NR	[redacted]
		North West	NR	NR	[redacted]
		South Central	NR	NR	[redacted]
		South East Coast	NR	NR	[redacted]
		South West	NR	NR	[redacted]
		West Midlands	NR	NR	[redacted]
		Yorkshire and the Humber	NR	NR	[redacted]
		Missing	NR	NR	[redacted]
		Type-2 Diabetes [% (N)]	49.8% (1856)	49.0% (1134)	[redacted]
		BMI, kg/m^2 [mean (SD)]	27.9 (5.4)	27.5 (5.1)	[redacted]
		BMI, kg/m^2 [% (N)] missing	NR	NR	[redacted]
		eGFR at baseline >= 60 ml/min/1.73m2 [% (N)]	51.8% (1931)	40.1% (929)	[redacted]
		eGFR at index, [% (N)] missing	NR	NR	[redacted]
		Prior hospitalisation for HF in 12 months prior [% (N)]	30.9% (1151)	29.2% (675)	[redacted]
		Prior atrial fibrillation or flutter [% (N)]	38.6% (1441)	46.5% (1076)	[redacted]
		Time since diagnosis (in years)			
		Mean, years (SD)	6.14 (6.32)	6.69 (6.62)	[redacted]
		0–1	18.6% (692)	15.8% (366)	NR
		1–5	37.9% (1415)	36.6% (848)	NR
		5+	43.5% (1623)	47.6% (1101)	NR

		Ischemic cause of HF [% (N)]	51.7% (1929)	57.1% (1322)	[redacted]
		LVEF [mean (%)]	26.8	30.2	[redacted]
		LVEF, [% (N)] missing	NR	NR	[redacted]
		Heart rate (bpm) [mean (SD)]	71.3 (11.7)	70.3 (11.5)	[redacted]
		Heart rate (bpm), [% (N)] missing	NR	NR	[redacted]
		NT-proBNP, pg/mL, [mean (SD)]	3034.7 (3665.5)	3226.9 (3788.5)	[redacted]
		NT-proBNP, pg/mL, median (IQR)	1910 (1115-3481)	2047 (SD)	[redacted]
		NT-proBNP, pg/mL, [% (N)] missing	NR	NR	[redacted]
		Blood pressure (mm Hg) [mean (SD)]	122.0 (15.6)	123.2 (15.7)	[redacted]
		Blood pressure (mm Hg), [% (N)] missing	NR	NR	[redacted]
		Implantable cardioverter-defibrillator [% (N)]	22.8% (851)	23.5% (543)	[redacted]
		Cardiac resynchronisation therapy [% (N)]	11.9% (442)	15.0% (347)	NR
		HF medication [% (N)]			
		ACEI or ARB + BB (no IVA, no ARNI)	62.7% (2339)	62.9% (1456)	[redacted]
		ARNI + BB (no IVA, no ACEI or ARB)	15.6% (580)	15.6% (360)	
		Other treatment	21.7% (811)	21.6% (499)	[redacted]
		EQ-5D score [mean (SD)]	0.7 (0.2)	0.7 (0.2)	NR

Footnote: a. As noted in Appendix O, the baseline characteristics are for the HFrEF-prevalent and incident population at index (2015); b. In PULSE, ischemic heart disease included myocardial infarction, coronary procedure and other IHD, cc. Primary care prescribing data only for PULSE – substantial underestimation expected
Abbreviations: NR, not reported; IHD, ischemic heart disease; DBP, diastolic blood pressure; SBP, systolic blood pressure

Key Issue 2: Uncertainty around the difference in efficacy of empagliflozin compared with standard of care in the Europe subgroup of EMPEROR-Reduced	**YES**	• **The baseline characteristics in the ITT and Europe subgroup in EMPEROR-R were broadly comparable, including the background use of ACEi or ARBs. Therefore, the Europe subgroup is unlikely to be a better representation of a typical UK HF patient than the ITT population.** • **As requested by the ERG, the results for all-cause-mortality, composite renal outcome and mean change from baseline in KCCQ-CSS score for the Europe subgroup are provided. Across these outcomes, empagliflozin demonstrated a benefit that was consistent with the ITT population, although the result was not statistically significant for KCCQ-CSS and all-cause mortality. These results are likely due to small sample and random variation rather than an underlying trend, as the Europe subgroup represented only 36% of the total trial population. Therefore, basing a recommendation for access for empagliflozin on the full ITT population would provide the Committee with more certainty.** • **A conclusion about the generalisability of the Europe subgroup of EMPEROR-R to UK clinical practice is hindered by their being a significant amount of missing data in some baseline characteristics in PULSE (e.g. BMI, ejection fraction, NT-proBNP, heart rate, eGFR), making it impossible to characterise a 'typical' UK HF patient.** • **Furthermore, the Europe subgroup is predominantly white (94.7%) and does not reflect the ethnic diversity in the UK, especially in metropolitan areas. The use of data from Europe subgroup to assess generalisability is not appropriate and could contribute to existing inequalities in access to healthcare (6), contrary to the NICE's Principles and the Equality Act 2010 (race is one of the protected characteristics) (2). This was an important consideration by the Committee in the appraisal of dapagliflozin (TA679) (7).**

		The ERG is concerned that there appears to be a reduction in efficacy with empagliflozin compared to placebo in the Europe geographical region subgroup analyses from EMPEROR-R. Therefore, the ERG requested the results for all-cause mortality, KCCQ-CSS and composite renal endpoint for the Europe subgroup. Like the >65 years subgroup, it is unlikely that the Europe subgroup is a better representation of a typical UK HF patient than the ITT population. The baseline characteristics for the ITT and Europe subgroup of EMPEROR-R were broadly similar (Table 2). In particular, the background ACE or ARB use (without ARNI) was 69.7% in the ITT population, 71.8% in the Europe subgroup compared to 73.3% in PULSE. The average age was also similar (69.5 years in Europe compared to 66.8 years in the ITT population). Across the Europe subgroup and the ITT population, the proportion who were male, the mean heart rate, mean systolic blood pressure, mean ejection fraction, mean BMI, proportion with T2DM status, NYHA, use of MRA and hospitalisations for heart failure in the previous 12 months was broadly similar. However, some differences were observed including the percentage with atrial fibrillation flutter and the percentage of patients with an eGFR ≥60ml/min.1.73m^2 at baseline. As requested by the ERG, the results for all-cause-mortality, composite renal outcome and mean change from baseline in KCCQ-CSS score for the Europe subgroup are reported below. All other trial outcomes for the Europe subgroup have already been provided during Clarification Questions. Across these outcomes, empagliflozin demonstrated a benefit that was consistent with the ITT population. For the renal composite outcome, this was statistically significant, however this was not the case for all-cause mortality and mean change from baseline in KCCQ-CSS. These results are likely due to small

		sample and random variation rather than an underlying trend, as the Europe subgroup represented only 36% of the total trial population. Therefore, basing a recommendation for access for empagliflozin on the full ITT population would provide the Committee with more certainty. • **Composite renal outcome** Like the ITT population, empagliflozin demonstrated a statistically significant benefit in composite renal outcome compared to PBO in the Europe subgroup. The benefit was even higher than in the ITT population [redacted] and HR[redacted] respectively) (Table 3, Figure 1). • **All-cause morality** The reduction in the risk of all-cause-mortality from empagliflozin compared to placebo was broadly similar in the ITT and Europe subgroup (8% and 12% reduction in the risk of all-cause mortality for empagliflozin vs placebo, respectively). The slightly higher reduction in the risk of all-cause-mortality in the Europe subgroup is likely due to low patient counts and is a spurious finding. In the Europe subgroup, at Day [redacted] of the empagliflozin group were at risk and 66/677 in the PBO group. Comparatively, in the ITT population at Day 810, [redacted] of the empagliflozin group and [redacted] of the PBO group and were still at risk (Table 4, Figure 2, Figure 3). • **Mean change from baseline in KCCQ-CSS** Like the ITT population, empagliflozin demonstrated a numerical improvement in the mean change from baseline to Week 12, 32 and 52 in the Europe subgroup compared to PBO. Although not statistically significant, this is likely due to underpowering from a small sample size. The analysis set for the KCCQ analysis in the Europe subgroup consisted of only 36% (1293/3529) of the full trial population (Table 5,

		Figure 4). A conclusion about the generalisability of EMPEROR-R to UK clinical practice is hindered by there being a significant amount of missing data of ~95% in some baseline characteristics in PULSE (e.g. BMI, ejection fraction, NT-proBNP, heart rate, eGFR). Some of these factors, such as NT-proBNP and eGFR are prognostic and help to identify patients who are at 'higher risk' for worse cardiovascular outcomes.(1) Missing data is an inherent limitation of any real-world evidence study, commonly encountered when using the CPRD as a data source. Therefore, it is impossible to define what is a typical HF patient in the UK and draw a definite conclusion about the generalisability of the EMPEROR-R trial population. Its likely that treatment improves, and awareness of HF improves over the next few years [as improving management is a key metric in the NHS Long Term Plan], it's likely that the age of diagnosis will decrease (8). Like the >65 years subgroup, use of the Europe subgroup represents an equity issue as it does not reflect the ethnic diversity in the UK. This was an important issue highlighted by the Committee during the appraisal for dapagliflozin in HFrEF (TA679) and was the key reason why the ITT population was considered the most generalisable to UK clinical practice (9). The Europe subgroup of EMPEROR-Reduced was 94.7% white compared to 70.5% in the ITT population. This can be compared to the multi-ethnic UK population, which consists of 86% white, 3.3% black, 7.5% Asian and 3.2% other (10). This difference is even wider in the metropolitan areas of the UK (44.9% white in London) (10). We considered whether other regions in EMPEROR-R reflected the ethnic diversity seen metropolitan areas in the UK. Of all the regions included, North America was the most ethnically diverse (70.8%

white) and was similar to the ITT population. These data suggest that the ITT population is more generalisable to the ethnically diverse UK population than the Europe subgroup and is, therefore, the population considered in the economic analysis for the original Company base case.

Table 2. Baseline characteristics of the ITT and Europe subgroup populations of EMPEROR-R

Variable	ITT (n = 3730) Combined	Europe (n = [redacted]	PULSE[a] (N= [redacted]
Age [mean (SD)]	66.8 (11.0)	[redacted]	[redacted]
Gender [% (N)]			
Male	76.1(2837)	[redacted]	[redacted]
Race [% (N)]		[redacted]	[redacted]
White	70.5% (2629)	[redacted]	[redacted]
Asian	18.0% (672)	[redacted]	[redacted]
Native	1.0% (39)	[redacted]	[redacted]
Black	6.9% (257)	[redacted]	[redacted]
Multiple/mixed	1.6% (61)	[redacted]	[redacted]
Pacific	0.4% (14)	[redacted]	[redacted]
Type-2 Diabetes [% (N)]	49.8 (1856)	[redacted]	[redacted]
BMI, kg/m^2, [mean (SD)]	27.9 (5.4)	[redacted]	[redacted]
eGFR at baseline >= 60 ml/min/1.73m2 [% (N)]	51.8% (1931)	[redacted]	[redacted]
eGFR at index, [% (N)] missing data	NR	[redacted]	[redacted]
Prior hospitalisation for HF in 12 months prior [% (N)]	30.9 (1151)	[redacted]	[redacted]
Prior atrial fibrillation or flutter [% (N)]	38.6 (1441)	[redacted]	[redacted]
Time since diagnosis (in years)			
Mean [SD)]	6.14 (6.32)	[redacted]	[redacted]
Median	3.99	[redacted]	[redacted]
0–1	18.6% (692)	[redacted]	[redacted]

		1–5	37.9% (1415)	[redacted]	[redacted]
		LVEF, [mean (SD)]	27.45 (6.03)		
		LVEF, [% (N)] missing data	NR		
		Heart rate, min−1, [mean (SD)]	71.3 (11.7)		
		Heart rate (bpm), [% (N)] missing data	NR		
		NT-proBNP, pg/mL, [mean (SD)]	3034.7 (3665.5)		
		NT-proBNP, pg/mL, median (IQR)	1910 (1115.0 to 3480.5)		
		NT-proBNP, pg/mL, [% (N)] missing data	NR		
		Blood pressure (mm Hg) [mean (SD)]	SBP: 122.0 (15.6)		
		Blood pressure (mm Hg), [% (N)] missing	NR		
		ACE or ARBs (no ARNI), [%,(N)]	69.7% (2600)		
		MRA, [%,(N)]	71.3% (2661)		
		Beta-blocker, [%,(N)]	94.7% (3533)		
		ARNI, [%,(N)], , [%,(N)]	19.5% (727)		
		Ivabradine, [%, (N)]	7% (260)		

Abbreviations, angiotensin receptor blocker; ARNI, angiotensin receptor neprilysin receptors; ACEi; angiotensin converting enzyme inhibitors; eGFR,estimated glomerular filtration rate; LVEF,left ventricular ejection fraction; NT-proBNP, N terminal pro b-type natriuretic peptide; IVA; ivabradine ; IQR, interquartile range; SD, standard deviation; NR, not recorded

a. The baseline characteristics are for the HFrEF-prevalent and incident population at index (2015)

Source data: EMPEROR-R baseline characteristics [combined and Europe subgroup]: Table 18.1.1.1.1; PULSE: Pulse report data on file

Table 3. Summary of the clinical endpoints for EMPEROR-R ITT and Europe Subgroup (composite renal outcome)

CLINICAL ENDPOINTS	EMPEROR-ITT Empagliflozin	EMPEROR-ITT placebo	EMPEROR-R Europe Empagliflozin	EMPEROR-R Europe placebo
Composite renal outcome[a]				
Patients with the composite renal event, N (%)	31 (1.6)	58 (3.1)	[redacted]	[redacted]
Incidence rate per 100 patient years at risk	1.56	3.07	[redacted]	[redacted]
Hazard ratio vs. placebo (95% CI), composite renal outcome	0.50 (0.32 - 0.77)		[redacted]	
Nominal p-value	0.0019		[redacted]	

a. Definition of composite renal outcomes: Time to the first event in the composite renal endpoint: chronic dialysis, renal transplant, or sustained reduction in eGFR (CKD-EPI)$_c$

		Figure 1. Estimated cumulative function for time to composite renal outcome in Europe subgroup of EMPEROR-R [redacted]

Table 4. Summary of the clinical endpoints for EMPEROR-R ITT and Europe Subgroup (all-cause mortality)

CLINICAL ENDPOINTS	EMPEROR-ITT Empagliflozin	EMPEROR-ITT placebo	EMPEROR-R Europe Empagliflozin	EMPEROR-R Europe placebo
All-cause mortality				
Number of patients with the event, N (%)	249 (13.4)	266 (14.2)	[redacted]	[redacted]
Incidence rate [patients with events per 100 patient years at risk]	10.17.6	10.78.1	[redacted]	[redacted]
Hazard ratio vs. placebo (95% CI), ACM	0.92 (0.77-1.10)		[redacted]	
Nominal p-value	0.35		[redacted]	

		Figure 2. All-cause Mortality KM plots for ITT randomised set of EMPEROR-R 

		Figure 3. All-cause Mortality KM plots for Europe subgroup 

Table 5. Mean change in KCCQ-CSS from baseline to Week 12, 32, and 52 (ITT and Europe subgroup)

Mean change from Baseline to:	ITT population			Europe subgroup		
	Empagliflozin (10mg)	Placebo	Diff	Empagliflozin (10mg)	Placebo	Diff
Baseline						
N	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mean (SE)	[redacted]	[redacted]		[redacted]	[redacted]	
Week 12[a]						
Adjusted Mean, SE, [95% CI], P[a]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 32[a]						
Adjusted Mean, SE, [95% CI], P[a]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 52[a]						
Adjusted Mean, SE, [95% CI], P[a]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]		[redacted]		[redacted]		

		Figure 4. Mean change from baseline to Week 52in KCCQ-CSS, OC-AD (Europe Subgroup) [redacted]

Key Issue 3: Uncertainty around the efficacy of empagliflozin compared with dapagliflozin	**YES**	• **There is limited evidence that dapagliflozin reflects standard of care, and therefore is not a relevant comparator for this appraisal. This is because the uptake is still significantly below the figures estimated in the NICE Resource Impact Template in TA679 (7)** • **A cost comparison case is the most appropriate decision framework because Bucher ITCs comparing empagliflozin and dapagliflozin across multiple outcomes showed [redacted], including the primary composite endpoint. Additionally, empagliflozin and dapagliflozin are priced at parity.** • **Prior technology appraisals and clinical guidelines suggest that this magnitude of uncertainty is acceptable. NICE committees have accepted a cost comparison case in previous appraisals where the 95% CI were similarly[redacted] as observed in these Bucher ITCs for dapagliflozin vs empagliflozin. The European Society for Cardiology (ESC) HF guidelines (11) recommended empagliflozin or dapagliflozin as a first line option in a broad range of patients with a Class 1A rating (i.e. gold standard), indicating that the clinical community also accept this uncertainty.** • **However, an incremental cost-effectiveness analysis has been provided upon the request of NICE and the ERG. The results showed that empagliflozin was [redacted] to dapagliflozin and offered a [redacted] of £[redacted] per patient, driven by an [redacted]. As expected, there was [redacted] in life years or QALYs. This is a conservative estimate as the CE model did not fully capture the relationship between renal decline and the**

		increased risk of HHF. Empagliflozin has demonstrated a numerical improvement in the slowing of renal function. • **When the difference in costs and benefits between the intervention and comparator are marginal, the ICER can be very sensitive. A scenario was undertaken where a mortality benefit in favour of dapagliflozin – based on the Bucher ITC estimate for all-cause and CV-mortality – was assumed for the first 23 months. A difference in QALY of only 0.03 resulted in an opposing conclusions, i.e empagliflozin is not cost effective vs is cost effective. This is seen in the probabilistic scatter plot where the probability of empagliflozin being cost effect was 59%. This further demonstrates that relying on a cost per QALY framework to base a recommendation for empagliflozin comes with significant risk. Therefore, a cost comparison case is the most appropriate.** • **Decision making should reflect the broader health system objectives and consistency. A recommendation comparable to dapagliflozin would support patient and clinician choice, support continuation of care for patients with comorbid diabetes and support the NHS's objective of reducing inequality in access to care by allowing broad prescribing of both dapagliflozin and empagliflozin across primary and secondary care.** The ERG considers the results of the Bucher ITC to be [redacted] and that the Company's assumption of [redacted] for empagliflozin and dapagliflozin lacks robustness. As a result, the ERG has requested an incremental analysis of empagliflozin vs dapagliflozin.

		A cost-utility comparison between empagliflozin and dapagliflozin is conceptually flawed as it only extrapolates the uncertainty in the Bucher indirect treatment comparison (ITC) of EMPEROR-Reduced and DAPA-HF trials. The Bucher ITC showed that there [redacted] for any of the outcomes tested (Table 6), including: o Primary endpoint: o composite endpoint of first adjudicated HHF or cardiovascular (CV) death. o Secondary endpoints: o time to first adjudicated HHF; o occurrence of adjudicated HHF (first and recurrent); o time to adjudicated CV death; o all-cause mortality; o worsening renal function; o change in KCCQ-TSS at 8 months. Both trials were high quality RCTs with large sample sizes, which were powered to detect a prespecified treatment effect on primary endpoint. The ERG concluded in the ERG report that both trials to have a low risk of bias. Thus, the most robust evidence on the relative efficacy of empagliflozin vs dapagliflozin comes from the Bucher ITC of the primary composite endpoint which suggests [redacted] (HR [redacted] (95% CI, [redacted]). The secondary endpoints were not powered to show a statistically significant effect and were not included in hierarchical testing in either trial (Table 6).

Table 6. Summary of Bucher ITC results for empagliflozin plus SoC versus dapagliflozin plus SoC

Endpoint: relative effect measure	EMPEROR-REDUCED: empagliflozin versus placebo[a]	DAPA-HF: dapagliflozin versus placebo[a]	Bucher ITC: empagliflozin versus dapagliflozin[a]
Primary endpoint			
Time to first event of adjudicated CV death or adjudicated HHF: HR (95% CI)	0.75 (0.65, 0.86)	0.75 (0.65, 0.85)	[redacted]
Secondary endpoint			
Time to first event of adjudicated CV death or adjudicated HHF (EMPEROR-Reduced) versus Time to first worsening of heart failure (hospitalisation or an urgent visit resulting in intravenous therapy for heart failure) or CV death (DAPA-HF) : HR (95% CI)	0.75 (0.65, 0.86)	0.74 (0.65, 0.85)	[redacted]
Time to first adjudicated HHF: HR (95% CI)	0.69 (0.59, 0.81)	0.70 (0.59, 0.83)	[redacted]
Time to adjudicated CV death: HR (95% CI)	0.92 (0.75, 1.12)	0.82 (0.69, 0.98)	[redacted]
Time to all-cause mortality: HR (95% CI)	0.92 (0.77, 1.1)	0.83 (0.71, 0.97)	[redacted]

		Occurrence of adjudicated HHF (first and recurrent) – analysed using a joint frailty model: HR (95% CI)	0.70 (0.58, 0.85)	0.71 (0.61, 0.82)	[redacted]
		Occurrence of adjudicated HHF (first and recurrent) – analysed using a Lin-Wei-Yang-Ying model: RR (95% CI)	0.76 (0.65, 0.89)	0.75 (0.65, 0.88)	[redacted]
		Worsening renal function (as defined in DAPA-HF): HR (95% CI)	0.52 (0.29, 0.92)	0.71 (0.44, 1.16)	[redacted]
		Change in KCCQ total symptom score at 8 months/7.4 months: MD (SE/95% CI)	1.6 (0.7)	2.8 (0.5)	[redacted]
		An incremental cost-effective analysis of health benefits and costs for the comparison of empagliflozin with dapagliflozin would depart from recent precedence in previous appraisals. In the appraisal of dapagliflozin for HFrEF (TA679), the Committee accepted equal efficacy between dapagliflozin vs sacubitril/valsartan given the uncertainty in the results of a Bucher ITC, whereby despite a trend favouring dapagliflozin, the 95% CI around the HR estimates for HHF and CV death similarly encompassed the no difference value of 1. In a recent appraisal of dapagliflozin in chronic kidney disease (ID3866), the Committee considered the outcomes for dapagliflozin and canagliflozin to be comparable (Committee meeting 14th October 2021), based on the result of ITCs, and was a minor topic for discussion. When DAPA-CKD was compared to CREDENCE, the HR's for the primary composite outcome and the secondary outcomes (all-cause mortality, end-stage renal disease, HHF) included 1, indicating no statistically significant difference. The Committee accepted the conclusion of equal efficacy. Similarly, in the appraisals for aflibercept (TA486) and guselkinumab (TA521), the			

		Committee acknowledged that uncertainty existed in the meta-analyses but still concluded that these technologies met the criteria for a cost comparison case. Therefore, based on precedence in previous appraisals, a cost comparison case of empagliflozin vs dapagliflozin is reasonable. There is limited evidence that dapagliflozin is standard of care, and therefore a relevant comparator for this appraisal. As stated in Doc B (Table 1), of all patients prescribed a HF medication in England, only 2.0% were prescribed dapagliflozin in the MQT August 2021. Although this has increased from 1.0% in MQT May 2021(12), it still far below the 20% estimated by NICE by the end of 2021 (7). The assumption of [redacted] and safety of these two SGLT2is is largely accepted by the clinical community based on the evidence from DAPA-HF and EMPEROR-Reduced trials, as indicated by: o The MHRA marketing authorisation for empagliflozin and dapagliflozin is identical; both products are indicated in adults for the treatment of symptomatic chronic heart failure with reduced ejection fraction (13, 14). o The ESC recently published updated guidelines of the management of HF. It recommended both dapagliflozin and empagliflozin in a broad population of adults with HFrEF with or without diabetes in line with the licensed indications (Supplementary Table 6(15)). The ESC guidelines states "Unless contraindicated or not tolerated, dapagliflozin or empagliflozin are recommended for all patients with HFrEF already treated with an ACEi/ARNI, a beta-blocker, and an MRA, regardless of presence or absence of diabetes (Section 5.3 and Supplementary Table 6(15)). This was a class 1A recommendation, indicating the highest quality of evidence.

		o Prominent UK and international HF experts consider that the optimal sequencing of treatments for HFrEF would entail initiating newly diagnosed HFrEF patients on a beta-blocker and an SGLT2i, with no distinction between molecular entities comprising this class of drugs (16). Such publicly available documentation suggests that the clinical community, including UK experts, are willing to accept a degree of uncertainty regarding the equivalence of trial populations and outcomes when prescribing empagliflozin and dapagliflozin in real-world HFrEF patients. Although BI believe that a cost comparison case is the most appropriate, an incremental cost-effective analysis of empagliflozin vs dapagliflozin was requested by the ERG. Consistent with ERG's preferred assumption in Issue 7 of no mortality benefit in the CE model for empagliflozin vs SoC, it is reasonable to assume the same for empagliflozin vs dapagliflozin. Like empagliflozin vs SoC, the empagliflozin and dapagliflozin KM CV-mortality curves [redacted] during the follow-up period of the trials and the HRs for the Bucher ITCs for all-cause and CV-related death were [redacted]. The difference in CV-mortality was [redacted] for months 0 to 11, a [redacted] was observed between months 13 to 20 before the KM curves began to [redacted] at month 23. When the PBO arms of EMPEROR-R and DAPA-HF were compared, the KM curves for CV-mortality followed [redacted] 13, 17 and 25, indicating no mortality benefit Figure 5). The all-cause mortality curve for dapagliflozin with follow-up after 24 months could not be accurately digitized, however since all-cause mortality and CV-mortality are related, a similar pattern can be expected. These observed trends mean that assuming [redacted] – based on a HR for CV-mortality and all-cause mortality, respectively, from the Bucher ITC - at each time point in the CE model is [redacted], and [redacted]. The magnitude of

		difference between empagliflozin vs PBO and empagliflozin vs dapagliflozin was [redacted]. When empagliflozin was compared to SoC, [redacted] difference in CV-mortality was observed compared to [redacted] when empagliflozin was compared to dapagliflozin ([redacted]). Therefore, the base case is that [redacted] between empagliflozin and dapagliflozin. Since the difference in costs and QALYs between empagliflozin and dapagliflozin was marginal, the result ICER is very sensitive. To characterise this uncertainty, a scenario was explored where dapagliflozin has a 12% mortality benefit for the first 23 months in the CE model, to align with the follow-up period in the trials before the KM CV-mortality began to converge (Figure ***5***). Although we consider this scenario analysis overly pessimistic with respect to empagliflozin it provides an opportunity to characterise the uncertainty based purely on the trial data.

		Figure 5. CV-mortality of empagliflozin vs dapagliflozin (naive comparison) 

		The base case incremental cost-effective analysis showed that empagliflozin was [redacted] to dapagliflozin and offered a [redacted] of [redacted] per patient; largely driven by [redacted]. As expected, there was [redacted] in life years or QALYs (**Figure 6**). The results demonstrate a [redacted] of probabilistic ICERs, indicative of the underlying uncertainty in the Bucher ITC, expanding across the [redacted] quadrant in which empagliflozin [redacted] (~[redacted] of iterations) to the [redacted] quadrant in which empagliflozin is [redacted] (~[redacted] of iterations). A similar trend was observed in a scenario analysis where a proportional CV-mortality benefit was assumed for the first 23 months. The ICERs spanned the SW and NE quadrants, with [redacted] of iterations being cost-effective in favour of empagliflozin (Figure 7, Table 8); with being dominant. These results demonstrate that the ICER is very sensitive for products with marginal and non-significant differences in efficacy that are priced at parity. This is illustrated in Table 9. The decision on whether empagliflozin is cost-effective is dependent on tiny differences in QALYs ranging from 0.001 to 0.05. A tiny difference in QALYs could mean that empagliflozin is not cost-effective. Basing a decision to recommended or not recommend empagliflozin in a similar position to dapagliflozin based on these marginal differences is a significant risk. Therefore BI believes that a cost comparison case is the most appropriate to consider for this comparison.

Table 7. Probabilistic incremental analysis of empagliflozin vs dapagliflozin (base case)

		Mean	Lower 95% CI	Upper 95% CI
Deterministic ICER	Costs	[redacted]	[redacted]	[redacted]
	QALYs			
	LYs			
	Cost per QALY			
Probabilistic ICER	Costs	[redacted]	[redacted]	[redacted]
	QALYs			
	LYs			
	Cost per QALY			

Abbreviations: NMB, Net Monetary Benefit, WTP, Willingness to Pay

Table 8. Probabilistic incremental analysis of empagliflozin vs dapagliflozin (scenario: 23-month mortality difference)

		Mean	Lower 95% CI	Upper 95% CI
Probabilistic ICER	Costs, difference	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	LYs			
	Cost per QALY			

Abbreviations: NMB, Net Monetary Benefit, WTP, Willingness to Pay

Table 9. Illustrative example of how marginal differences in QALYs impacts the ICER

	Difference in QALYs (empagliflozin vs dapagliflozin)	-0.01	-0.001	-0.05
Assuming a cost saving for empagliflozin of [redacted], as per base case (PSA	Saving's per QALY lost	[redacted]	[redacted]	[redacted]
	Likely decision			

Abbreviations: NMB, Net Monetary Benefit, WTP, Willingness to Pay

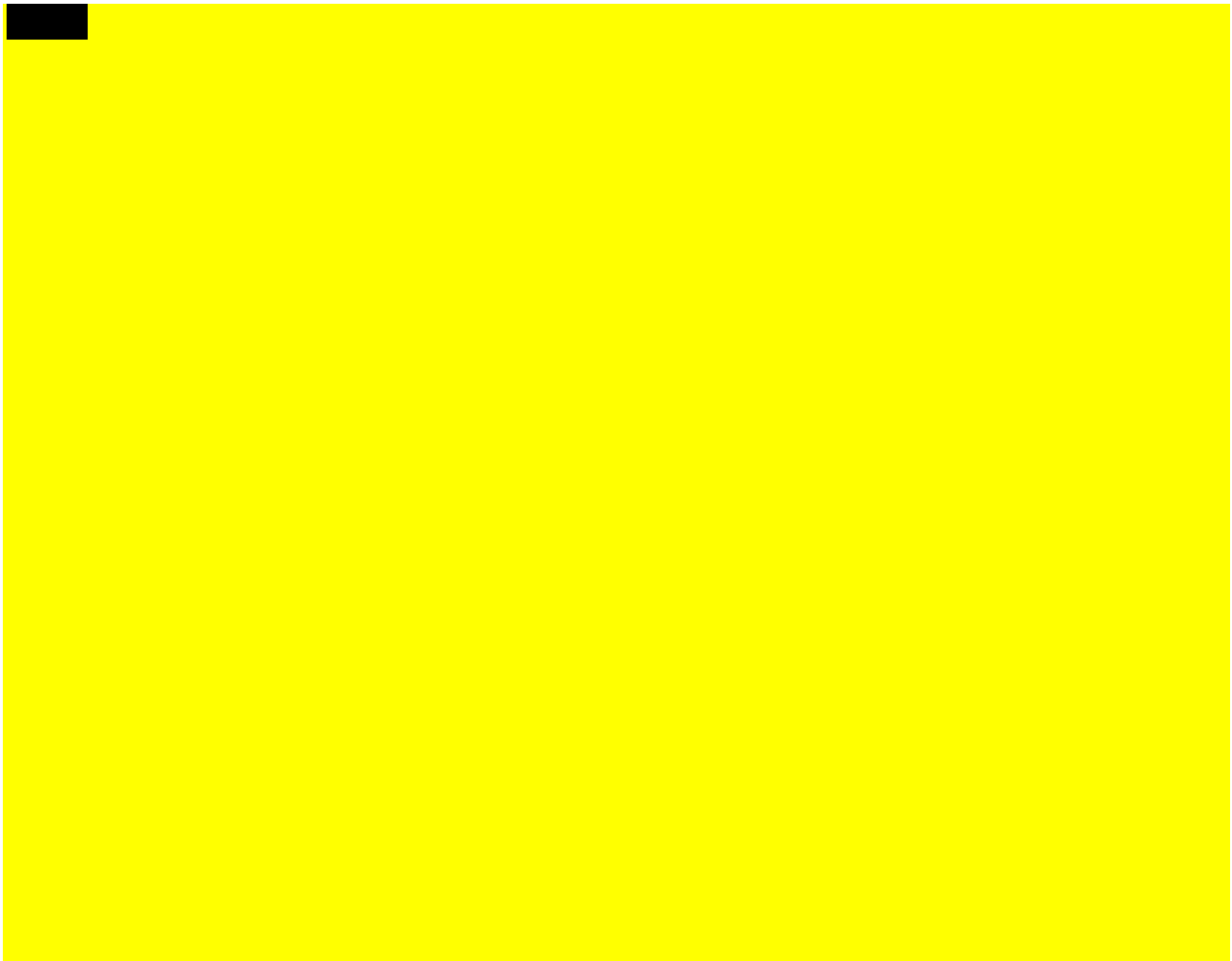
		Figure 6. Probabilistic scatter plot showing the distribution of ICERs for empagliflozin vs dapagliflozin (no mortality difference through the time horizon of the model) 

		Figure 7. Probabilistic scatter plot showing the distribution of cost per QALY for empagliflozin vs dapagliflozin (scenario: 23 months benefit) 

		The incremental cost-effective estimate for empagliflozin vs dapagliflozin is [redacted] because it did not fully capture the relationship between the decline in renal function and increased risk of HHF. This was done to simplify the modelling approach. Empagliflozin demonstrated [redacted] decline in renal function than dapagliflozin (Bucher ITC: HR for empagliflozin vs dapagliflozin in worsening in renal function: [redacted]), (Table 6). Data from a nested UK study of 50,114 HF patients indicated that these relationships are clinically important. Lawson et al 2018 reported in that the prevalence of CKD (eGFR<60 ml/min/1.73m^2) in the HF community was 63%, and was associated with a 11% increase in hospitalisation risk (17). Therefore, basing a recommendation solely on the results of an incremental cost-effectiveness estimate represents a missed opportunity to improve clinically important outcomes for patients. Decision making should also reflect the broader health system objectives and consistency. A recommendation comparable to dapagliflozin would support: • patient and clinician choice; • continuation of care for patients with comorbid diabetes; • the clinical community belief in equivalence of SGLT2 inhibitors; • support the NHS's objective of reducing inequality in access to care by allowing broad prescribing of both dapagliflozin and empagliflozin across primary and secondary care; • and would exemplify consistency in NICE Committee's decision making for treatments with similar efficacy and identical cost.

Key Issue 4: The modelling of patients' distribution across the KCCQ-CSS health states	**YES**	The ERG requested that the Company: 1. ***Clarify which dataset from EMPEROR-R is being used to estimate the TPs*** The transition probabilities were derived using KCCQ-CSS measurements from the observed case including data after treatment discontinuation (OC-AD) dataset. Data from baseline and weeks 12, 32 and 52 visits were used for analyses with and without imputation using last observation carried forward (LOCF) for visits where measurements were not observed for patients who were still followed prior to the visit. Imputation did not have a major impact on transition probabilities. 2. **Provide the data from EMPEROR-R that allowed the estimation of TPs and proportion of patients in each KCCQ-CSS in the model** The observed transition probabilities in the two arms of the trial are provided in Table 10 & Table 11 These represent movements between quartiles over the baseline to Week 12, Week 12 to 32, and Week 32 to 52 periods as observed in the trial. These are converted to monthly probabilities by finding the p^{th} root of the matrices. As noted by the ERG, this is only possible when the observed transition probabilities over the three periods form a non-singular matrix. The ERG also noted: "*As the Company did not share the original matrices, the ERG has no way of confirming the non-singularity of the matrices and the appropriateness of the method used.*" We confirm the observed matrices were non-singular, as, otherwise, the procedure to derive the roots would have failed and not yielded any results. The proportion of patients in each KCCQ-CSS quartile in the model can be seen on the Markov Trace sheet of the model. The distribution of patients across quartiles at a given time can be derived by normalizing the proportions at that time (i.e., dividing each by the

sum across the fourth quartiles). These results were previously provided in Table 35 of responses to clarifications.

Table 10. Observed Transition Probability Matrices after imputation by LOCF for KCCQ-CSS Health States in the Empagliflozin 10 mg Arm

From/To	(0, 55.2)	(55.2,75)	(75,89.6)	(89.6,100]
Baseline to Week 12 (3-month transition probability)				
(0, 55.2)	[redacted]	[redacted]	[redacted]	[redacted]
(55.2,75)	[redacted]	[redacted]	[redacted]	[redacted]
(75,89.6)	[redacted]	[redacted]	[redacted]	[redacted]
(89.6,100]	[redacted]	[redacted]	[redacted]	[redacted]
Week 12 to Week 32 (5-month transition probability)				
(0, 55.2)	[redacted]	[redacted]	[redacted]	[redacted]
(55.2,75)	[redacted]	[redacted]	[redacted]	[redacted]
(75,89.6)	[redacted]	[redacted]	[redacted]	[redacted]
(89.6,100]	[redacted]	[redacted]	[redacted]	[redacted]
Week 32 to Week 52 (5-month transition probability)				
(0, 55.2)	[redacted]	[redacted]	[redacted]	[redacted]
(55.2,75)	[redacted]	[redacted]	[redacted]	[redacted]
(75,89.6)	[redacted]	[redacted]	[redacted]	[redacted]
(89.6,100]	[redacted]	[redacted]	[redacted]	[redacted]

Table 11. Observed Transition Probability Matrices after imputation by LOCF for KCCQ-CSS Health States in the Placebo arm

From/To	(0, 55.2)	(55.2,75)	(75,89.6)	(89.6,100]
Baseline to Week 12 (3-month transition probability)				
(0, 55.2)	[redacted]	[redacted]	[redacted]	[redacted]
(55.2,75)	[redacted]	[redacted]	[redacted]	[redacted]
(75,89.6)	[redacted]	[redacted]	[redacted]	[redacted]
(89.6,100]	[redacted]	[redacted]	[redacted]	[redacted]
Week 12 to Week 32 (5-month transition probability)				
(0, 55.2)	[redacted]	[redacted]	[redacted]	[redacted]
(55.2,75)	[redacted]	[redacted]	[redacted]	[redacted]
(75,89.6)	[redacted]	[redacted]	[redacted]	[redacted]
(89.6,100]	[redacted]	[redacted]	[redacted]	[redacted]
Week 32 to Week 52 (5-month transition probability)				
(0, 55.2)	[redacted]	[redacted]	[redacted]	[redacted]
(55.2,75)	[redacted]	[redacted]	[redacted]	[redacted]
(75,89.6)	[redacted]	[redacted]	[redacted]	[redacted]
(89.6,100]	[redacted]	[redacted]	[redacted]	[redacted]

		3. Produce the TPs observed in EMPEROR-R for the KCCQ-CSS quartiles defined in the model and explain how these relate to the mean changes reported in the trial. The use of KCCQ-CSS quartiles as health states in the model makes it impossible to relate the distribution of patients in quartiles in the model to mean changes in KCCQ-CSS scores over time reported in the trial. Relating the two by assessing based on mean scores at baseline and mean changes between visits to predict the probability or timing of movements between quartiles (as described on page 101 of the ERG report) is not appropriate as these means do not reflect the variability in individual changes that would drive movements between quartiles. For example, while the mean change from baseline to Week 12 was 5.19 with empagliflozin and 3.25 with placebo (per page 101 of ERG report), 25% of patients had improved by more than 12 points with empagliflozin and 11 points with placebo, and 10% had decreased by more 11 points with empagliflozin and 14 points with placebo (Table 12, Table 11). –Agreement between the model-derived changes in KCCQ-CSS and those observed in the trial can only be made reliably based on comparison of the distribution of patients across KCCQ-CSS quartiles over time observed in the trial vs. the model. -This was previously provided in Table 35 of responses the clarification question document (Question B12). The ERG noted not being able to reproduce the model results. The reported distributions are simply calculated as the normalized (i.e., divided by total) proportion of patients in each KCCQ-CSS quartile (total on and off treatment) taken from the empagliflozin and SoC Markov tables in the back-end sheets of the model at months 0, 3, 8, and 12.

Table 12. Distribution of Individual Changes in KCCQ-CSS from Baseline to W12 visit (without Imputation) Treatment Arm

	N	Mean	SD	Min	P10	P25	Median	P75	P90	Max
Empagliflozin 10mg	[redacted]									
Placebo	[redacted]									

4. **Conduct scenario analyses where the effect of empagliflozin seen at month 8 in the model (sustained by the combination of the proportion of patients in the better KCCQ-CSS states in the empagliflozin arm at month 8 and the low probability of disease progression for both SoC and empagliflozin arms in month 9+) wanes over time.**

   In the Company base case, it is assumed that the benefit of empagliflozin on KCCQ is maintained from the end of the trial period (Week 52) for the remaining duration of a patient's life. Evidence suggests that this is a reasonable assumption. The EMPA-REG study – a CVOT trial of T2DM patients with multiple CV risk factors –showed that all-cause mortality, CV-mortality and heart failure hospitalisation was sustained for 3.1 years (Shown in Figure 1 of reference (18)). Although KCCQ was not collected as part of the EMPA-REG study, it demonstrated a sustained consistent effect for empagliflozin over time. This assumption is consistent with prior NICE appraisals for products in the same drug class. In the dapagliflozin appraisal in HFrEF (TA679), the Committee concluded that "*there was no*

		evidence for or against treatment waning in the long-term. Clinical experts and stakeholders confirmed that treatment with dapagliflozin would likely be lifelong". A scenario was built in the CE the model (see Context tab rows 70-72) to test the impact of a treatment effect loss on the results, as it was specifically requested by the ERG. In this scenario, two assumptions were tested: 1) At a defined point in time the proportion of patients in the KCCQ-CSS quartiles under the treatment arm was set equal to those proportions in the placebo arm at 5, 3, 2, and 1 years. a. For the trial population, the ICER increased from £4,717 to £4,813, £4,935£5,054, and £5,232 once 5-year, 3-year, 2-year, and 1-year, time points were tested, respectively. 2) The transition probabilities (TPs) between KCCQ-CSS quartiles for treatment arm were set to the TPs for the SoC arm after 8 months. a. This scenario increased the ICER from £4,717 to £5,688. Thus, even if waning was included, the impact on the cost-effectiveness of empagliflozin vs SoC was marginal and below the £20,000 willingness to pay threshold.

Key Issue 5: Use of a Poisson model to estimate hospitalisation for heart failure	YES	• **The base case allowed the rates to increase as patients progressed over time to worse KCCQ-CSS health states. This approach is consistent with the accepted approach in previous appraisals [TA267 [ivabradine)(19), TA388 (sacubitril valsartan) (20), TA679 (dapagliflozin)(7)].** • **The alternative suggested by the ERG, which is to use time to event data from EMPEROR-R to also vary the HHF rate within each KCCQ-CSS health state over time in the CE model yielded counter-intuitive results. The analyses showed a declining rate in hospitalisation over time.** • **Clinical experts confirmed that they would not expect the rate of HHF to decline over time, but rather increase. The observed patterns are not due to a deficiency of the method but driven more likely by the declining numbers of patients at risk and events near the end of follow-up in the trial. Therefore, this analysis cannot be reliably used in the CE model as it would introduce more uncertainty.** • **To reflect clinical reality, a scenario option on the Context tab of the model (rows 61-64) where a constant increase in the rate of HHF for both the empagliflozin and SoC arms within each KCCQ-CSS state was added. Results from this scenario suggested that allowing the rate of HHF to increase over time reduces the ICER due to higher reduction in number of HHFs (ICER reduces from £4,717 to £4,492 in the example provided in the model). This had no meaningful impact on the ICER.**

		The ERG remains uncertain if HHFs are accurately estimated in the long-term, despite the CE model's ability to accurately reproduce the number of HHF observed in the 18-month trial period. This is because, consistent with prior appraisals (TA267 [ivabradine](19), TA388 (sacubitril valsartan) (20), TA679 (dapagliflozin)(7)), the Company assumed that the overall rate of HHF remained constant over the lifetime of the CE model. Nonetheless, the ERG recommended that the Company undertook a scenario analysis where HHF KM data from EMPEROR-R is used to model HHF rates over the time horizon in the CE model. The ERG has proposed to use "HHF KM data" to model time to hospitalisation using parametric survival models to project over time to overcome reservations about the constant rate assumption invoked in the Poisson model currently applied in the model. It should be noted that parametric survival models are not typically applied for recurrent events; variations in the rate of event over time can be handled in the context of event count models (like Poisson) by allowing time as a predictor to allow for variations and provides similar flexibility as the parametric approach. We have provided results from Poisson models with time and log-time as predictors previously (Clarification Question B8). These showed negative slopes for time, suggesting declining rates of hospitalisation over time, which is not clinically plausible and cannot be used reliably in the economic model. Clinical experts confirmed that they would not expect the rate of HHF to decline over time. The observed patterns are not due to a deficiency of the method but driven more likely by the declining numbers of patients at risk and events near the end of follow-up in the trial. To examine this further, we implemented the proposed parametric approach as this effectively allows testing a broader range of temporal trends in the hazards of HHF events. The data were arranged using a counting process setup with start and stop times to create periods defined by the occurrence of each hospitalisation

		and/or changes in KCCQ-CSS quartiles. That is, a patient will have one record per change in KCCQ-CSS and per hospitalisation with start and stop times of the period defined by the time when these changes occur. Parametric models were then fitted to these data with *flexsurvreg* in R (a package in r to fit KM data to parametric survival distributions) testing exponential, Weibull, Gompertz, log-logistic, log-normal and generalised gamma distributions. Typically, such analyses would include a random-effect or frailty to account for repeated records in the data on the same individual; this option is not available currently in *flexsurvreg* and could not be used. The hazards produced by these fits are shown in Figure 8 (dashed lines are monthly observed rates per person-month in the groups). All the tested distributions yielded fits with decreasing or plateauing hazards as seen with the Poisson model with time as predictor, and thus, lead to similarly implausible long-term projections of declining risk if applied in the model (Figure 8). These findings support the idea that the fitted Poisson model – which is equivalent to the exponential parametric form – represents the most clinically plausible fit to the data while allowing the rate to increase as patients progress over time through inclusion of time-varying KCCQ-CSS as a predictor. The assumption of a constant hazard of the long-term may not hold, and it is possible that rates may increase over time as patients progress and age, but this increasing pattern could not be derived from the trial data and is, therefore, more appropriately examined through scenario analyses in the model. A scenario option is added on the Context tab of the model (rows 61-64) where a constant increase in the rate of HHF can be examined by KCCQ-CSS state. Results from this scenario suggested that allowing the rate of HHF to increase over time reduces the ICER due to higher reduction in number of HHFs (ICER reduces from £4,717 to £4,492 in the example provided in the model).

		Figure 8. Parametric fit of HHF rates in EMPEROR-R (first and recurrent) 

Key Issue 6: Overestimation of hospitalisation for heart failure in the UK population	**YES**	• **The original Company base case, which used HHF rate data from the ITT population in EMPEROR-R, is the most relevant to consider.** • **As stated in Issue 1, use of the >65 years subgroup introduces uncertainty into the analysis as it includes only 63% of the EMPEROR-R data. Further, using the >65 years subgroup as the base case represents an equity in access issue, as younger HF patients tend to belong to a lower socio-economic or ethnic minority group.** • **Although differences in HHF and CV-mortality rates were observed between EMPEROR-R and PULSE, this is likely due to inaccurate reporting than clinically meaningful differences in patient characteristics. A comparison of all-cause mortality, which is an objective outcome not affected by recording differences suggests a more similar risk profile between the populations.** • **Although BI believe the original base case is the most appropriate, at the ERG's request a scenario was run in the model where the HHF for empagliflozin and SoC was adjusted to reflect the lower HHF rates observed in PULSE. This resulted increased ICER from £4,717 to £7,000 QALY gained (empagliflozin vs SoC), which is still below the £20,000 willingness to pay threshold.** The ERG was concerned that the HHF rates were lower in PULSE than in EMPEROR-R and requested that the Company undertook a scenario analysis where the "HF KM data" for the >65 years subgroup is used to model time to HHF in the CE model and adjustments made to reflect the HHF rates observed in PULSE. The difference in HHF and CV-mortality rates observed in EMPEROR-R and PULSE is likely due inaccurate recording of events in PULSE rather than differences in patient characteristics. Unlike in the real-world, HHF and CV-mortality in EMPEROR-R were adjudicated by Committee according to a strict protocol. In the real-world, an

		elderly patient might be admitted to wards other than cardiology, and therefore HHF and CV-mortality may not be recorded as the primary reason for hospitalisation because general physicians and other specialists may not recognise the symptoms of acute HF. Further evidence that the PULSE dataset is unreliable is the substantial amount of missing data (Table 1) and that the largest cohort in PULSE was patient 'unknown' [redacted]), indicating no definite diagnosis of HFrEF. Limitations in CPRD data have recently been highlighted in a recent clinical audit of medical records and SNOMED CT coding for 78 GP practices (864194 population) in the UK for HF. Of 19 393 patients' records that were audited, the HF case finder identified 9725 additional patients to be audited, of whom 2916 patients with HFrEF required further codes (47% increase) (21). Nonetheless, at the ERG's request, an analysis exploring whether time to event data for HHF in the >65 years could be implemented in the CE model to reflect the PULSE HHF rates was undertaken. Given the limitations described in the previous paragraph, firm conclusions cannot be made about the generalisability of EMPEROR-R to UK clinical practice based on these analyses. Table 13 summarises the observed rates in PULSE (full population, HFrEF subgroup and prevalent subset of the HFrEF subgroup) and the placebo arm from EMPEROR-R (full population and the 65 years and older subgroup).

Table 13. Observed rates of events in PULSE and EMPEROR-R Placebo groups

Group	N	PYs	HHF	HHF per 100PY	ACM	ACM per 100PY	CVM	CVM per 100PY
PULSE-ALL	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
HFrEF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
HFrEF-Prevalent	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Placebo	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Placebo - 65+	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

As noted in responses to Issue 5, the use of parametric modelling for recurrent HHF events does not offer an improvement over the Poisson modelling framework and is retained as the approach in the model. Switching to a parametric approach is not necessary to adjust predicted rates to reflect lower event counts as in PULSE. This can be achieved by calibrating the predicted rates from the current model in each cycle by an appropriate factor to achieve a targeted overall rate. For example, an adjustment factor that would calibrate the model to the rate observed in PULSE can be obtained as the relative rate observed in PULSE and that observed among placebo patients in EMPEROR-R from a joint regression model of the individual patient-level data from the two sources and including a term in the model for the population source (PULSE vs. EMPEROR-Placebo). This was done with a negative binomial model for the total number of HHF events observed for patients with follow-up duration as an offset (log-transformed). A negative binomial model was used as this was done in the original analyses of the PULSE data to address overdispersion in event counts. This yielded a rate-ratio of 0.43 when comparing patients in PULSE with HF with reduced ejection fraction (HFrEF) and patients aged 65 years or older from the placebo arm in EMPEROR-R, and 0.44 when considering only patients with prevalent HF at index in the HFrEF subgroup.

		A comparison of all-cause mortality, which is an objective outcome not affected by diagnostic differences, suggests a more similar risk profile between the populations. The all-cause mortality rate in PULSE was [redacted] per 100 person-years (PY) compared with 10.7 per 100PY with placebo in EMPEROR (12.3 per 100PY in the 65+ subgroup). The rate was [redacted] per 100PY among patients with prevalent HFrEF. Quantifying the relative difference in mortality in a Cox model fitted to the joint data for the PULSE and placebo populations was 1.02, and 0.76 when comparing the prevalent HFrEF and placebo 65+ subgroups. This disproportionate difference between HHF and overall mortality may be in part be due to the impact of different assessment methods for HHF. Thus, calibrating the model to the observed HHF rates in PULSE may not accurately reflect outcome in the UK population. Nevertheless, the model was run with a HHF rate-ratio of 0.43 using a scenario option added to the Context tab (row 66) of the CE model to calibrate to the PULSE rate of HHF. The rate-ratio lowered the overall number of hospitalisations and as a result increased ICER from £4,717 to £7,000 QALY gained (empagliflozin vs SoC)

Key Issue 7: Modelling of mortality	YES	• **The ERG considered that there was not enough evidence to support the inclusion of a treatment effect in CV and non-CV-mortality in the economic model as the KM all-cause mortality curves hardly separated over the follow -up period of EMPEROR-R. The Company concluded that *"while the curves are not exactly parallel for the treatment arms, the deviations are largely due to the crossing of the curves in the tail, which is likely caused by small patient counts".*** • **Given the ERGs comments about the uncertainty in the death benefit for empagliflozin, the Company has adopted a new base case, where no direct death benefit is assumed; aligned with the ERG position. However, in EMPEROR-R, patients in the empagliflozin arm had a higher probability of remaining in a better KCCQ-CSS states over time compared with SoC patients. Being in a better KCCQ-CSS health state is associated with a lower probability of death, and therefore it is reasonable to maintain this indirect death benefit in the CE model. This results in an ICER of £4,999/QALY compared to £4,717 in the original company base case. (Please see "Summary of changes to the Company's cost-effectiveness estimate(s)").** • **Even in the most conservative scenario when no mortality benefit is assumed for empagliflozin compared to SoC either directly or indirectly through being in a different KCCQ-CSS health state, cost-effectiveness is demonstrated and did not vary significantly from the original Company base case (£4,777/QALY vs £4,717/QALY, respectively).** The ERG conducted a scenario analysis where CV and non-CV-mortality were assumed to be the same in the empagliflozin and the SoC arms. The ERG noted that when no treatment effect is assumed for empagliflozin on mortality in the economic model, there is still a benefit associated with empagliflozin on both CV and non-CV-mortality. This is because the probability of patients dying is different in every KCCQ-CSS state of the model. Given that patients in the empagliflozin arm of the model have a higher probability of remaining in the better

KCCQ-CSS states over time compared with SoC patients, the former also experience a lower probability of death. The ERG recommended that the Company consider adding a scenario analysis in the model where it is assumed that empagliflozin has no survival benefit over SoC (including through the residency in KCCQ-CSS states).

A scenario was implemented into the CE model where it was assumed that empagliflozin has no survival benefit over SoC and no survival benefit could be captured through residency in KCCQ-CSS states (see Context tab row 68). A Weibull function was fitted to the data from the placebo arm of the EMPEROR-R trial for ACM and CVM with no KCCQ predictors (

Table ***14*** and

Table ***15***) for ACM and CVM functions for the ITT and 65+ subgroups, respectively). These two functions were applied in both arms in the model to capture mortality and resulted in minimal impact on ICER (£4,777 vs £4,717).

Table 14.Scenario analysis: No survival benefit for empagliflozin and no survival benefit captured through KCCQ-CSS health state residency (ITT population)

ITT Population					
All-cause mortality					
Distribution	Parameter	Estimate	SE	Lower 95% CI	Upper 95% CI
WeibullPH	shape	[redacted]	[redacted]	[redacted]	[redacted]
WeibullPH	scale	[redacted]	[redacted]	[redacted]	[redacted]
CV-related mortality					

Distribution	Parameter	Estimate	SE	Lower 95% CI	Upper 95% CI
WeibullPH	shape	[redacted]	[redacted]	[redacted]	[redacted]
WeibullPH	scale	[redacted]	[redacted]	[redacted]	[redacted]

Table 15. Scenario analysis: No survival benefit for empagliflozin and no survival benefit captured through KCCQ-CSS health state residency (>65 years subgroup)

>65 years subgroup					
All-cause mortality					
Distribution	Parameter	Estimate	SE	Lower 95% CI	Upper 95% CI
WeibullPH	shape	[redacted]	[redacted]	[redacted]	[redacted]
WeibullPH	scale	[redacted]	[redacted]	[redacted]	[redacted]
CV-related mortality					
Distribution	Parameter	Estimate	SE	Lower 95% CI	Upper 95% CI
WeibullPH	shape	[redacted]	[redacted]	[redacted]	[redacted]
WeibullPH	scale	[redacted]	[redacted]	[redacted]	[redacted]

Key Issue 8: Overestimation of mortality in the UK population	**YES**	• **The ERG noted differences in the number of CV-deaths and all-cause deaths observed in PULSE vs the Company's model for the ≥65 years subgroup (Table 43, ERG report).** • **Although the differences in outcomes could be due to EMPEROR-R being enriched with a 'sicker' population, this conclusion is very uncertain. The significant amount of missing data observed in the baseline characteristics in PULSE makes it impossible to characterise a typical UK HF patient.** • **The strongest evidence for this is that the all-cause mortality across PULSE and EMPEROR-R was comparable, as noted in Issue 6, with a hazard ratio of 1.02 when compared in a Cox analysis. All-cause mortality is an objective measure that is not subject to any reporting bias.** • **When an adjustment to CV-mortality and all-cause mortality to reflect the PULSE HFrEF (prevalent and incident patients at index) was applied to the >65 years subgroup in the CE model, the length of patient's life was extended life-years and resulted in higher reduction in number of HHFs. This in turn, resulted in a decrease in the ICER from £4,717 to £4,325/QALY gained.** The ERG noted that when compared to PULSE, the SoC arm of the model overestimated the number of CV-related deaths (and underestimated the number of non-CV-deaths) when the >65 years subgroup from EMPEROR-R was modelled. The ERG concluded that this is because the EMPEROR-R population was more likely to die of CV causes than PULSE patients. During clarification questions, the Company conducted a scenario analysis where the data for the subgroup analysis from EMPEROR-R for patients above 65 years was used to estimate cost-effectiveness (Clarification Question B5). The ERG requested that the KM data for all-cause mortality and CV-mortality in the >65 years subgroup is supplied.

		As described in issue 1 and 6, the difference in HHF and CV-mortality rates observed in EMPEROR-R and PULSE is likely due inaccurate recording of events in PULSE rather than a clinically meaningful difference in the patient characteristics. Although the differences in outcomes could be due to EMPEROR-R being enriched with a 'sicker' population, this conclusion is very uncertain. The significant amount of missing data observed in the baseline characteristics in PULSE makes it impossible to characterise a typical UK HF patient. Due this this inaccurate recording in PULSE, any adjustment in the model to reflect the mortality observed in PULSE is uncertain. The strongest evidence for this is that the all-cause mortality across PULSE and EMPEROR-R was comparable, as noted in Issue 6, with a hazard ratio of 1.02 when compared in a Cox analysis (Table 16). A similar approach as the one described in Issue 6 was used to derive adjustment factors to calibrate mortality in the CE model. The factors were derived as hazard ratios from joint analyses of the patient-level data from PULSE and EMPEROR-R placebo patients contrasting all-cause and cardiovascular (CV) mortality with a Cox model with a term for population. The estimated hazard ratios represent the relative differences in mortality and multiplying hazards (or exponentiating survival probabilities) in the CE model by this factor yields an adjusted survival aligning with that observed in PULSE. Analyses were based on the HFrEF subgroup in PULSE and the 65+ subgroup in the placebo arm from EMPEROR, as well as considering the prevalent subset of the HFrEF subgroup as sensitivity analysis, and the ITT population (Table 16).

Table 16 Hazard ratios comparing mortality in PULSE vs. EMPEROR-R placebo populations.

	All-Cause Mortality	CV-Mortality
HFrEF vs. Placebo (65+)	[redacted]	[redacted]
Prevalent HFrEF vs. Placebo (65+)	[redacted]	[redacted]
PULSE vs. Placebo	[redacted]	[redacted]

The proportionality of the relative difference in AC-mortality and CV-mortality between the PULSE HFrEF group (incident and prevalent) and the >65 years group in EMPEROR-R was assessed using the log-log survival vs. log-time plots (Figure 9 and Figure 10). While the curves cross very early in the follow-up (<0.5 years), the lines remain relatively parallel, particularly for CV-mortality, suggesting use of a constant HR is acceptable.

		Figure 9. Log-log survival vs log-time for AC- mortality for PULSE vs EMPEROR-R >65 years subgroup (naive comparison) [redacted]

		Figure 10. Log-log survival vs log-time for CV- mortality for PULSE vs EMPEROR-R >65 years subgroup (naive comparison) [redacted] As discussed in similar analyses for HHF (Issue 6), comparisons of CV-mortality may be affected by limitations around complete identification of HF-related events in PULSE. Nevertheless, the HRs for HFrEF vs. Placebo (65+) were applied in the model to calibrate to the PULSE survival using the scenario option added on the context tab (rows 76-77). This scenario extended patients' life-years and resulted in higher reduction in number of HHFs which resulted in lowering ICER from £4,717 to £4,325/QALY gained. The ERG also requested KM curves for mortality in EMPEROR-R for the 65+ subgroup. This is provided in Figure 11 and Figure ***12***.

		Figure 11. KM curves for CV-mortality comparing empagliflozin vs placebo in the >65 years subgroup in EMPEROR-R [redacted]

		Figure 12. KM curves for AC-mortality comparing empagliflozin vs placebo in the >65 years subgroup in EMPEROR-R 

Key Issue 9: Impact of hospitalisation for heart failure in patients' quality of life	**YES**	• **The ERG expressed concerns that the disutility per HHF assumed that the impact on HRQoL of HHF in the CE model lasted for 1 year.** • **The Company clarifies that the utility equation includes indicators for time since the hospitalisation rather than duration.** The ERG considered that the impact of HHF on patients' quality of life is overestimated in the CE model. The Company estimated the -0.246 disutility per HHF event by assuming that all HHFs in the model last for 1 year. The ERG considers it unlikely that all HHFs events will have that duration. The ERG advised that the Company reported the proportion of patients in EMPEROR-R who were hospitalised for 1; 2; and 8 months and generate a weighted disutility value to be applied in the CE model. Ideally, for the UK population, the same analysis would be conducted using PULSE data, as the mean duration of HHF is likely to be lower in PULSE than in EMPEROR-R. The Company clarifies that the utility equation includes indicators for time since the hospitalisation rather than duration. More specifically, the equation includes terms for 0-1, 1-2, 2-4, 4-12 months from hospitalisation vs. not hospitalised ever or in past 12 months. The coefficient for 0-1 represents the change in utility in the first month after hospitalisation, 1-2 represents the change in the second months after hospitalisation, etc (shown in the CE model in Risk Equations – Lookup!D10-13). These describe the course of change in utilities over the year following the hospitalisation with patients returning to their pre-hospitalisation after one year. Therefore, length of hospitalisations does not play into the calculations in any way.

Key Issue 10: Quality of life regressions for the UK population	**YES**	• **During clarification questions, the cost-effectiveness of the >65 years subgroup was estimated using a utility regression equation for the ITT population but with a binary predictor for >65 years to adjust for the impact of age. This resulted in an ICER for £6,342/QALY.** • **The utility regression equation was re-estimated using only the >65 years subgroup rather than the ITT population.** • **This re-analysis increased the ICER slightly to £6,362/QALY for the 65+ subgroup.** The baseline characteristics from the >65 years EMPEROR-R subgroup were used in the QoL regression analysis; however the regression was not re-estimated in this subgroup and thus the coefficients for the predictors remained the same as those for the ITT population. The ERG recommended that the Company re-estimate the regression model using the subgroup data. The regression analyses for utilities were repeated in the subgroup of patients aged 65 years or older. The coefficients from equations from the ITT population and 65+ subgroup are described in Table 17. Using effects derived from the subgroup equation in the economic model, the ICER slightly changed from the base case value of £6,342 to £6,362 for the 65+ subgroup.

Table 17. Health-related quality of life equation derived from EMPEROR-Reduced trial (>65 years subgroup)

	Equation Based ITT Population			Equation Based on 65+ Subgroup		
Predictors	**Estimate**	**SE**	**p-value**	**Estimate**	**SE**	**p-value**
Intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
EQ-5D-3L at baseline (standardised)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Updated KCCQ-CSS quartile (Ref.: ([redacted]))						
(55.2,75)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
(75,89.6)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[89.6,100]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
hHF events in the prior year (Ref.: No hHF Event/Past 12 months)						
hHF in the prior 0–1 month: 'Yes'	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
hHF in the prior 1–2 month: 'Yes'	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
hHF in the prior 2–4 month: 'Yes'	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
hHF in the prior 4–12 month: 'Yes'	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Treatment-emergent AE in the prior month (Ref. No AE/Past 1 month)						
HI	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
UTIs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GIs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Acute RF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
HG	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Age>=65 (Ref.: Age<65)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Male (Ref.: Female)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Region (Ref.: Europe)						
Latin America	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
North America	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Asia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Other	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

		Ischemic HF: 'Yes' (Ref.: 'No') [redacted]
Key Issue 11: Sex distribution underlying utility estimates	**YES**	• **The utility value, 0.7740. applied to KCCCQ-CSS health state 4 was taken from Sullivan et al. This value was based on a UK dataset where 52% were female.** • **This differed from the gender distribution in EMPEROR-R and PULSE (24% females and 35% females, respectively).** • **However, when the 0.7740 utility value was adjusted to reflect the gender balance in EMPEROR-R and PULSE, utility increased slightly to 0.7831 and 0.7795, respectively.** The ERG was concerned that the utility value associated with the KCCQ-CSS health state did not reflect the gender distribution in EMPEROR-R or PULSE. The ERG suggested that the Company undertake the following analyses: • Trial population: o Adjusts the 0.7740 value in the trial population analysis to reflect the gender distribution in EMPEROR-R. • UK population: o Adjusts the 0. 7740 value in the trial population analysis to reflect the gender distribution in PULSE. The estimated EQ-5D utility values for the KCCQ-CSS health state were 0.7831 for EMPEROR-REDUCED and 0.7795 for PULSE after adjusting for gender (***Figure 13***). These utility values were higher than the Sullivan et al. 2011(22) study because the proportion of females is lower in each case. Updating the utility value associated with the KCCQ-CSS quartile 4 state in the model to be based on gender

distribution in EMPEROR-REDUCED or PULSE had minimal impact on utility.

Figure 13. Utility adjustment for gender for KCCQ-CSS 4 health state

	Sex	%	EQ-5D overall
Sullivan et al 2011	Female	52.0	0.8150
	Male	48.0	0.850
	Ratio, M/F	1.04 [0.85/0.815]	

→

EQ-5D [60-69 years]	0.7740
Estimated EQ-5D, Female, 60-69	0.7584
Estimated EQ-5D, Male, 60-69	0.7909

	Sex	%	EQ-5D overall (Estimated)
EMPEROR-R	Female	24	0.7831
	Male	76	
PULSE	Female	[redacted]	[redacted]
	Male	[redacted]	

Key Issue 12: Quality of life gains in EMPEROR-Reduced	**YES**	• **When Issue 4, 6, 8 9,10 and 11 were combined to create the most pessimistic scenario, it only reduced the QALY benefit from 0.20 in the original Company base case to 0.13. the resulting ICER was £7,123/QALY gained, comfortably below the £20,000/QALY willingness to pay threshold.** The ERG is concerned that the EQ-5D data from EMPEROR-R [redacted] in empagliflozin patients' QoL when compared to placebo patients and that the economic model generates a QALY gain of [redacted]. The two main drivers of QALY gain in the model are related to: 1) how much longer empagliflozin patients stay in the better KCCQ-CSS states; and 2) the reduction in HHF experienced by empagliflozin patients. The ERG requested the following analyses: For the trial population: • The scenario analysis suggested for the KCCQ-CSS modelling (see Issue 4) in combination with the adjustments to the QALY calculations suggested in Issue 9 and Issue 11. For the UK population: • A combination of the following scenario analyses: The scenario analysis suggested for the KCCQ-CSS modelling (see Issue 4); the adjustments to the QALY calculations suggested in Issue 9, Issue 10 and Issue 11; the scenario analysis suggested in Issue 6 to reduce the number of HHF in the model; the scenario analysis suggested in Issue 8 to reduce the number of CV-deaths in the model These analyses are reported in Table 18.

Table 18. Scenario analysis: adjustment to utility values

For trial population		
#	Scenario	Quality of life gain
I	Base case	0.20
II	scenario analysis suggested for the KCCQ-CSS modelling (see Issue 4) in combination with the adjustments to the QALY calculations suggested in Issue 9 and Issue 11 (sudden treatment effect loss at 3 years; utility value associated with the KCCQ-CSS quartile 4 state is set to 0.7831)	0.13
For UK population		
#	Scenario	Quality of life gain
I	Base case	0.20
II	Scenario analysis suggested for the KCCQ-CSS modelling (see Issue 4) (Based on sudden treatment effect loss at 3 years)	0.14
III	When applying the adjustments to the QALY calculations suggested in Issue 9, Issue 10 and Issue 11 (utility value associated with the KCCQ-CSS quartile 4 state is set to 0.7795)	0.20
IV	scenario analysis suggested in Issue 6 to reduce the number of HHF in the model	0.18
V	scenario analysis suggested in Issue 8 to reduce the number of CV-deaths in the model	0.21
VI	#II, #III, #IV, #V	0.13

Section 1.6	**YES**	Given the difficulties around validating the model KCCQ-CSS results against the trial, discussed in Issue 4, the ERG recommended that the company provides additional data from EMPEROR-R in the format outlined in Table 16. The ERG notes that the company should complete Table 16 twice; once for the ITT population in EMPEROR-R, and another for the ≥65 years subgroup from the trial. The summary of event counts by KCCQ-CSS quartiles over time are provided in the tables below for the EMPEROR-R ITT population (OC-AD) as well as for the subset of patients aged 65 years or over. Periods are defined as the first 3 months of follow-up (0 to 91.32 days [3*30.44]), 4 to 8 months (92 to 243.52 days) and beyond (> 243.52 days). Patients are grouped based on observed or imputed KCCQ-CSS levels at baseline, week 12 and week 32, which correspond to the start of each period; events are counted in periods based on the time of the event and the patient's KCCQ-CSS quartile at the start of the period in which the event occurs. For example, suppose a patient in Q3 at baseline has an event before week 12, at which time their KCCQ-CSS is in Q2; the event would be counted under Q3 in the Month 1-3 period. A direct comparison of these counts with results from the model for validation may not be reliable, however, since the summaries below count events based on *prior* KCCQ-CSS levels whereas the model applies transition probabilities at each cycle and counts events in their *current* KCCQ-CSS level. For instance, in the above example, the patient may have progressed to Q2 prior to the event in the model, in which case the event would count under Q2. Thus, any comparisons between the summaries and the model must be interpreted carefully.

Table 16A. KCCQ-CSS data from EMPEROR-R for model validation

Data	Empagliflozin + SoC			Placebo + SoC		
	Months 1-3	Months 4-8	Months 9+	Months 1-3	Months 4-8	Months 9+
KCCQ 1						
n/N*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of HHF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of CV-deaths	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of all-cause deaths	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ 2						
n/N*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of HHF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of CV-deaths	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of all-cause deaths	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ 3						
n/N*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of HHF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Number of CV-deaths	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of all-cause deaths	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ 4	[redacted]					
n/N*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of HHF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of CV-deaths	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of all-cause deaths	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: CV, cardiovascular; EMPEROR-R, EMPEROR-Reduced; HHF, hospitalisation for heart failure; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire clinical summary score.

*n/N – please provide the number of patients in the respective KCCQ-CSS state divided by the total number of patients alive at the same time point

Table 16B. KCCQ-CSS data from EMPEROR-R for model validation among patients aged 65 years and over.

Data	Empagliflozin + SoC			Placebo + SoC		
	Months 1-3	Months 4-8	Months 9+	Months 1-3	Months 4-8	Months 9+
KCCQ 1						
n/N*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of HHF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of CV-deaths	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of all-cause deaths	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ 2	[redacted]					
n/N*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of HHF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of CV-deaths	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of all-cause deaths	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KCCQ 3						
n/N*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of HHF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of CV-deaths	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

		Number of all-cause deaths	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		KCCQ 4	[redacted]					
		n/N*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Number of HHF	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Number of CV-deaths	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Number of all-cause deaths	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Abbreviations: CV, cardiovascular; EMPEROR-R, EMPEROR-Reduced; HHF, hospitalisation for heart failure; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire clinical summary score. *n/N – please provide the number of patients in the respective KCCQ-CSS state divided by the total number of patients alive at the same time point						

Please use the table below to respond to additional issues in the ERG report that have not been identified as key issues. Please do **not** use this table to repeat issues or comments that have been raised at an earlier point in this appraisal (e.g. at the clarification stage).

Issue from the ERG report	Relevant section(s) and/or page(s)	Does this response contain new evidence, data or analyses?	Response
None			

NICE National Institute for Health and Care Excellence

Summary of changes to the Company's cost-effectiveness estimate(s)

Company:

If you have made changes to the Company's preferred cost-effectiveness estimate(s) in response to technical engagement, please complete the table below to summarise these changes.

Key Issue(s) in the ERG report that the change relates to	Company's base case before technical engagement	Change(s) made in response to technical engagement	Impact on the Company's base case ICER
Key Issue 7: **Modelling of mortality**	**A small mortality benefit was considered for empagliflozin vs. SoC based on analysis of EMPEROR-R trial data**	**No mortality benefit was assumed for empagliflozin vs. SoC. The treatment predictor in the ACM and CVM equations was set to zero; however a mortality benefit is derived indirectly from being in a particularly KCCQ-CSS health state as empagliflozin demonstrated a statistically significant benefit in mean change in KCCQ-CSS from baseline.**	**For the trial population, ICER increased from £4,717 to £4,999 (6% increase)** **For the UK population, ICER increased from £6,342 to £7,270 (15% increase)**
Issue 3: Comparison of dapagliflozin vs empagliflozin	**As the Bucher ITC suggests outcomes for empagliflozin and dapagliflozin are [redacted] e, a cost comparison case was presented**	**Implemented an incremental cost-effective analysis for empagliflozin vs dapagliflozin**	**Empagliflozin was [redacted] compared to dapagliflozin, and offered a [redacted] of [redacted] per patient; driven by an increased slowing of decline in renal function**

References

1. Spinar J, Spinarova L, Malek F, Ludka O, Krejci J, Ostadal P, et al. Prognostic value of NT-proBNP added to clinical parameters to predict two-year prognosis of chronic heart failure patients with mid-range and reduced ejection fraction - A report from FAR NHL prospective registry. PLoS One. 2019;14(3):e0214363.
2. NICE. Our Priniciples: Social Value Judgements 2021 [Available from: https://www.nice.org.uk/about/who-we-are/our-principles.
3. Conrad N, Judge A, Tran J, Mohseni H, Hedgecott D, Crespillo AP, et al. Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. The Lancet. 2018;391(10120):572-80.
4. Lawson Claire A, Zaccardi F, Squire I, Okhai H, Davies M, Huang W, et al. Risk Factors for Heart Failure. Circulation: Heart Failure. 2020;13(2):e006472.
5. Government U. Government to publish Levelling Up White Paper 2021 [Available from: https://www.gov.uk/government/news/government-to-publish-levelling-up-white-paper.
6. Public Health England. Local action on health inequalities Understanding and reducing ethnic inequalities in health. . 2018.
7. NICE. Dapagliflozin for treating chronic heart failure with reduced ejection fraction [TA 679] 2021 [Available from: https://www.nice.org.uk/guidance/ta679.
8. NHS. NHS Long Term Plan 2019 [Available from: https://www.longtermplan.nhs.uk/publication/nhs-long-term-plan/.
9. National Institute for Health and Care Excellence. Final appraisal document: Dapagliflozin for treating chronic heart failure with reduced ejection fraction. 2020.
10. Gov.uk. Population of England and Wales 2020 [Available from: https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/population-of-england-and-wales/latest.
11. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Bohm M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J. 2021;42(36):3599-726.
12. IQVIA. LPD, IQVIA Ltd, incorporating data derived from THIN, A Cegedim Database, May 2021.
13. Summary of Product Characteristics. Jardiance 2020 [Available from: https://www.ema.europa.eu/en/documents/product-information/jardiance-epar-product-information_en.pdf.
14. www.medicines.org.uk. Summary of Product Characteristics - Dapagliflozin Last updated: August 2021 [

16. McMurray JP, M. How Should We Sequence the Treatments for Heart Failure and a Reduced Ejection Fraction? Circulation. 2021;143(9):875-7.
17. Lawson CA, Testani JM, Mamas M, Damman K, Jones PW, Teece L, et al. Chronic kidney disease, worsening renal function and outcomes in a heart failure community setting: A UK national study. International journal of cardiology. 2018;267:120-7.
18. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, et al. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. N Engl J Med. 2015;373(22):2117-28.

19. National Institute for Health and Care Excellence. Ivabradine for treating chronic heart failure [TA267] 2012 [Available from: https://www.nice.org.uk/guidance/ta267.
20. National Institute for Health and Care Excellence. Sacubitril valsartan for treating symptomatic chronic heart failure with reduced ejection fraction [TA388] 2016 [Available from: https://www.nice.org.uk/guidance/ta388.
21. Kahn M, Grayson AD, Chaggar PS, Ng Kam Chuen MJ, Scott A, Hughes C, et al. Primary care heart failure service identifies a missed cohort of heart failure patients with reduced ejection fraction. Eur Heart J. 2021.
22. Sullivan PW, Slejko JF, Sculpher MJ, Ghushchyan V. Catalogue of EQ-5D scores for the United Kingdom. Med Decis Making. 2011;31(6):800-4.

NICE National Institute for Health and Care Excellence

Patient expert statement and technical engagement response form

Empagliflozin for treating chronic heart failure with reduced ejection fraction [ID3826]

Thank you for agreeing to give us your views on empagliflozin and its possible use in the NHS.

You can provide a unique perspective on conditions and their treatment that is not typically available from other sources.

About this Form

In **part 1** we are asking you to complete questions about living with or caring for a patient with the condition.

In **part 2** we are asking you to give your views on key issues in the Evidence Review Group (ERG) report that are likely to be discussed by the committee. An overview of the key issues are summarised in the executive summary at the beginning of the ERG report.

The key issues in the ERG report reflect the areas where there is uncertainty in the evidence, and because of this the cost effectiveness of the treatment is also uncertain. In part 2 of this form we have included any of the issues raised by the ERG where we think having a patient perspective could help either:

- resolve any uncertainty that has been identified
  or
- provide missing or additional information that could help committee reach a collaborative decision in the face of uncertainty that cannot be resolved.
-

In **part 3** we are asking you to provide 5 summary sentences on the main points contained in this document.

If you have any questions or need help with completing this form please email the public involvement team via pip@nice.org.uk (please include the ID number of your appraisal in any correspondence to the PIP team).

Please return this form by **5pm** on **21 October 2021**

Completing this form

Part 1 can be completed anytime. We advise that the final draft of part 2 is completed after the expert engagement teleconference (if you are attending/have attended). This teleconference will briefly summarise the key issues, any specific questions we would like you to answer and the type of information the committee would find useful.

Please use this questionnaire with our hints and tips for patient experts. You can also refer to the Patient Organisation submission guide. **You do not have to answer every question** – they are prompts to guide you. There is also an opportunity to raise issues that are important to patients that you think have been missed and want to bring to the attention of the committee. The text boxes will expand as you type.

Important information on completing this expert statement

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you want to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 15 pages.

PART 1 – Living with or caring for a patient with chronic heart failure with reduced ejection fraction and current treatment options	
About you	
1.Your name	**Nick Hartshorne-Evans**
2. Are you (please tick all that apply):	☒ a patient with chronic heart failure with reduced ejection fraction ? ☐ a patient with experience of the treatment being evaluated? ☐ a carer of a patient with chronic heart failure with reduced ejection fraction ? ☒ a patient organisation employee or volunteer? ☐ other (please specify):
3. Name of your nominating organisation.	Pumping Marvellous Foundation
4. Has your nominating organisation provided a submission? Please tick all options that apply.	☐ No, (please review all the questions below and provide answers where possible) ☒ Yes, my nominating organisation has provided a submission ☐ I agree with it and **do not wish to** complete a patient expert statement ☒ Yes, I authored / was a contributor to my nominating organisations submission ☐ I agree with it and **do not wish to** complete this statement ☒ I agree with it and **will be** completing

5. How did you gather the information included in your statement? (please tick all that apply)	☒ I am drawing from personal experience. ☒ I have other relevant knowledge/experience (e.g. I am drawing on others' experiences). Please specify what other experience: We run the largest peer to peer support network in the UK for people living with heart failure. ☐ I have completed part 2 of the statement **after attending** the expert engagement teleconference ☒ I have completed part 2 of the statement **but was not able to attend** the expert engagement teleconference ☐ I have not completed part 2 of the statement
Living with the condition	
6. What is your experience of living with chronic heart failure with reduced ejection fraction ? If you are a carer (for someone with chronic heart failure with reduced ejection fraction) please share your experience of caring for them.	I was diagnosed in 2010 with heart failure and have lived with HF since then. I am the founder and CEO of the UK's Heart Failure charity. I am considered a key opinion leader and patient expert in patient insights, advocacy and helping people live with the condition. I have worked with NICE for approximately 7 years in the field of HF and get involved in STA's, MTA's, Guidelines and Scientific Projects. I work with NHS England as a patient expert advisor. I am also the patient expert on the NHS Clinical Entrepreneur Programme Board.
Current treatment of the condition in the NHS	
7a. What do you think of the current treatments and care available for chronic heart failure with reduced ejection fraction on the NHS?	Having been involved with all the current guideline creation and appraisals of new technologies for the past 7 years, I believe that if guidelines are followed and STA decisions are followed and offered to patients who are clinically appropriate for specific treatments then the suite of medications is clearly defined. Heart failure treatments need to constantly evolve therefore I am an advocate of evidenced

7b. How do your views on these current treatments compare to those of other people that you may be aware of?	continuation of 7a - based, cost-effective treatments increasing in prevalence as choice is important like it is in other LTC's e.g., Cancer and Diabetes amongst others. 7b – My views represent the opinion of our community of the 10,000's of HF patients and families we represent along with the 920,000+ people with heart failure.
8. If there are disadvantages for patients of **current NHS treatments** for chronic heart failure with reduced ejection fraction (for example how the treatment is given or taken, side effects of treatment etc) please describe these	There is always a fine balance between the prognostic value and cost effectiveness that the system sees vrs the impact on QOL for people living with the condition and at the hard end of the treatments available. There are very few options available to people living with heart failure in terms of drug class and their effects. The more options available, gives the prescriber clinical choice which is important due to the increasing individualised treatment plans that patients are on. First and second line therapies in heart failure, predominantly in the chronic state are oral medication. The quantity of tablets taken can be overpowering for people with heart failure, especially if several co-morbidities are included. Side-effects are common at first and tend to subside which enables prescribers to up-titrate and optimise, which can be a difficult journey for people living with HF. Choice is paramount as we learn how to treat heart failure whilst ensuring other conditions are suitably considered. The vast majority of patients would welcome choice as it gives them hope that there is a therapy that will suit them and help them live better with heart failure.
Advantages of this treatment	
9a. If there are advantages of empagliflozin over current treatments on the NHS please describe these. For example, the impact on your Quality of Life your	From a patient perspective it seems that Empagliflozin has the same benefit as another recent addition to heart failure medications, Dapagliflozin. Feedback from our community on Dapagliflozin has been, on the whole very positive with few side effects, mostly tolerated well. Although it may be an adjunct to the algorithm in the NG106 CHF guidelines for adults 2018 patients are being prescribed it. Advantages

ability to continue work, education, self-care, and care for others? 9b. If you have stated more than one advantage, which one(s) do you consider to be the most important, and why? 9c. Does empagliflozin help to overcome/address any of the listed disadvantages of current treatment that you have described in question 8? If so, please describe these.	• Choice in a drug class that has demonstrated efficacy and cost-effectiveness It demonstrates efficacy and this is the most important aspect for patients especially around their QOL which enables them, downstream, to lead a more productive life. People don't want heart failure, but we can support them to live better with it. Empagliflozin is an important adjunct to the current set of treatments. 1. It gives choice to both the prescribing clinician and to the patient ref eGFR 2. It gives HOPE to patients there is another evidenced based medication the NHS is adopting to beat HF. 3. As the treatment options are increased and improved, it builds resilience into the supply chain that the patient and the system are not just reliant on a single treatment in a specific drug class (SGLT2i) 4. It is a one tablet treatment and well tolerated.
Disadvantages of empagliflozin	
10. If there are disadvantages of empagliflozin over current treatments on the NHS please describe these? For example, are there any risks with empagliflozin? If you are concerned about any potential side affects you have heard about, please describe them and explain why.	I have no concerns or see any disadvantages in Empagliflozin for the treatment of just heart failure and or heart failure with T2DM. A small number patients who have started SGLT2i's comment on urinary tract infections, these either subside or stop when the SGLT2i is removed.

Patient population	
11. Are there any groups of patients who might benefit more from empagliflozin or any who may benefit less? If so, please describe them and explain why. Consider, for example, if patients also have other health conditions (for example difficulties with mobility, dexterity or cognitive impairments) that affect the suitability of different treatments	There may be groups of patients who are clinically not suitable for Empagliflozin therefore unable to benefit. These patients will probably not be suitable for the whole drug class. As the amount of tablets, a person with heart failure needs to take on a daily basis is generally significant, adding in Empagliflozin shouldn't be a cause for concern for the patient, or in terms of medication adherence from the prescribing clinician. Due to potential increased urination, this may cause initial inconvenience or potentially discomfort but most patients with be on an oral diuretic so they should be used to this challenge.
Equality	
12. Are there any potential equality issues that should be taken into account when considering chronic heart failure with reduced ejection fraction and empagliflozin? Please explain if you think any groups of people with chronic heart failure with reduced ejection fraction are particularly disadvantaged. Equality legislation includes people of a particular age, disability, gender reassignment, marriage and	Just like a recent addition to the NHS, Dapagliflozin in HFrEF, I am concerned that it is made clear and simple to understand that if Empagliflozin is given its marketing authority, it is prescribed, not just from a specialist position, but also by Primary Care GPs on the advice of a heart failure specialist. I also believe that communication needs to be stronger around the general use of and SGLT2i's in Primary Care with patients who have heart failure, with or without T2DM. There is no reason to withhold Empagliflozin based on age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex and sexual orientation or people with any other shared characteristics.

civil partnership, pregnancy and maternity, race, religion or belief, sex, and sexual orientation or people with any other shared characteristics More information on how NICE deals with equalities issues can be found in the NICE equality scheme More general information about the Equality Act can and equalities issues can be found at https://www.gov.uk/government/publications/easy-read-the-equality-act-making-equality-real and https://www.gov.uk/discrimination-your-rights.	
Other issues	
13. Are there any other issues that you would like the committee to consider?	I want NICE to consider that there are many benefits to adding new, innovative treatments to heart failure other than cost-effectiveness. As we come out of the pandemic, we need to ensure that we at least offer the most optimised therapy regime for people living with heart failure. It is the ultimate responsibility of NICE to ensure crystal clear dissemination of how to enable this and who is responsible for prescribing.

PART 2 – Technical engagement questions for patient experts	
Issues arising from technical engagement	
We welcome your response to the questions below, but you do not have to answer every question. If you think an issue that is important to patients has been missed in the ERG report, please also advise on this in the space provided at the end of this section. The text boxes will expand as you type. Your responses to the following issues will be considered by the committee and may be summarised and presented in slides at the appraisal committee meeting. For information: the patient organisation that nominated you has been sent a technical engagement response form (a separate document) which asks for comments on each of the key issues that have been raised in the ERG report, these will also be considered by the committee.	
14: Please provide any responses to Key Issue 1: Uncertainty around the generalisability of the results from EMPEROR-Reduced to the older heart failure with reduced ejection fraction population expected in clinical practice (EMPEROR-Reduced included people with more severe heart failure with reduced ejection	A third of the recruited cohort were from Europe where clinical practice generally reflects the practice of the NHS. Although the trial population was younger and varied in a degree around ethnicity the results of the trial reflected the UK HF population and standards of care.

fraction than expected in clinical practice)	
15. Please provide any responses to Key Issue 2: Uncertainty around the difference in efficacy of empagliflozin compared with SoC in the Europe subgroup of EMPEROR-Reduced (The ERG has requested complete baseline characteristics for the Europe subgroup of EMPEROR-Reduced to explore differences in efficacy in the Europe geographical region subgroup analyses)	
16. Please provide any responses to Key Issue 3:	

Uncertainty around the efficacy of empagliflozin compared with dapagliflozin (The company carried out an indirect treatment comparison to compare empagliflozin with dapagliflozin but ERG considers the method used is uncertain)	
17. Please provide any responses to Key Issue 4: The modelling of patients' distribution across the KCCQ-CSS health states (ERG suggest it is difficult to validate modelled health states against the trial and have suggested the company provide additional analyses)	

18: Please provide any responses to Key Issue 5: Use of a Poisson model to estimate hospitalisation for heart failure (ERG suggest time to hospitalisation for heart failure should be modelled using Kaplan-Meier data from Emperor-Reduced)	
19. Please provide any responses to Key Issue 6: Overestimation of hospitalisation for heart failure in the UK population (ERG suggests the 3-year outcomes in the company model overestimate the number of hospitalisations for heart failure in the standard of	

care arm. It suggests time to hospitalisation for heart failure in the UK population should be modelled using Kaplan Meier data from the above 65 years subgroup in EMPEROR-Reduced)	
20. Please provide any responses to Key Issue 7: Modelling of mortality (ERG have suggested an alternative approach to modelling all cause and cardiovascular related mortality)	
21. Please provide any responses to Key Issue 8: Overestimation of mortality in the UK population	

(ERG suggest the standard of care arm of the model overestimates the number of cardiovascular-related deaths (and underestimates the number of non-cardiovascular deaths. It has suggested the company provide Kaplan Meier data from EMPEROR-Reduced for all-cause and CV related mortality in the above 65 years subgroup)	
22. Please provide any responses to Key Issue 9: Impact of hospitalisation for heart failure in patients' quality of life (ERG suggest company assumptions overestimate the impact of hospitalisations for	

heart failure on people's quality of life. It suggests an alternative approach to be applied in the model)	
23. Please provide any responses to Key Issue 10: Quality of life regressions for the UK population (The company Utility scores were analysed using mixed-effects linear regression using all available EQ-5D measurements across all visits but ERG suggest this should be re-estimated for the UK population subgroup data)	
24. Please provide any responses to Key Issue 11:	

Sex distribution underlying utility estimates (ERG suggests the company adjusts the health state utility values generated to reflect the baseline gender distribution more accurately in EMPEROR-Reduced)	
25. Please provide any responses to Key Issue 12: Quality of life gains in EMPEROR-Reduced (ERG suggest the company adjust its approach to modelling EQ-5D to accurately reflect the quality of life gains in EMPEROR- Reduced)	

26. Are there any important issues that have been missed in the ERG report?	

PART 3 -Key messages

1. In up to 5 sentences, please summarise the key messages of your statement:

- Empagliflozin significantly improves QOL measures in HFrEF vrs the placebo arm.
- Empagliflozin is easily initiated with no titration, well tolerated, manageable side effects and an easy addition to SOC.
- The medication class of SGLT2i is starting to demonstrate that it is a significant treatment and potentially powerful addition to SOC. SGLT2i's should be considered as SOC with more Data in the future. NICE should consider a revision of the NG106 guidelines, published in September 2018, focussing on the treatment algorithm.
- Empagliflozin demonstrated in EMPEROR-HF, a trial with UK relatability, reduced mortality, and hospitalisation rates.
- It should not be underestimated the growing problem of heart failure in the UK. The pandemic has accelerated the urgency to better manage HF. Patients under the care of the NHS need optimised treatment options and treatments over and above SOC as outlined by NG106 which was 2017-2018 in the making prior to SGLT2i consideration. It is important for patients that this happens as they all want a better quality of life, to live longer and reduced visits whether planned or unplanned to health system entrance points.

Thank you for your time.

Please log in to your NICE Docs account to upload your completed statement, declaration of interest form and consent form.

.....

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

……

NICE National Institute for Health and Care Excellence

Technical engagement response form

Empagliflozin for treating chronic heart failure with reduced ejection fraction [ID3826]

As a stakeholder you have been invited to comment on the ERG report for this appraisal. The ERG report and stakeholders' responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the key issues below. You do not have to provide a response to every issue. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be included in the committee papers in full and may also be summarised and presented in slides at the appraisal committee meeting.

Deadline for comments by **5pm** on **21 October 2021**

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the ERG report which summarises the background and submitted evidence, and presents the ERG's summary of key issues, critique of the evidence and exploratory analyses. This will provide context and describe the questions below in greater detail.
- Please ensure your response clearly identifies the issue numbers that have been used in the executive summary of the ERG report. If you would like to comment on issues in the ERG report that have not been identified as key issues, you can do so in the 'Additional issues' section.
- If you are the company involved in this appraisal, please complete the 'Summary of changes to the company's cost-effectiveness estimates(s)' section if your response includes changes to your cost-effectiveness evidence.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response unreadable. Please type information directly into the form.
- Do not include medical information about yourself or another person that could identify you or the other person.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.
- If you provide journal articles to support your comments, you must have copyright clearance for these articles.

- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow, and all information submitted under 'depersonalised data' in pink. If confidential information is submitted, please also send a second version of your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received during engagement are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

About you

Your name	[redacted]
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	AstraZeneca UK Ltd
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	None

Key issues for engagement

Please use the table below to respond to questions raised in the ERG report on key issues. You may also provide additional comments on the key issue that you would like to raise but which do not address the specific questions.

Key issue	Does this response contain new evidence, data or analyses?	Response
Key issue 1: **Uncertainty around the generalisability of the results from EMPEROR-Reduced to the older heart failure with reduced ejection fraction population expected in clinical practice**	**NO**	We agree with the ERG that there is uncertainty around the generalisability of the EMPEROR-R trial to the UK population. In particular, the EMPEROR-R trial has a more severe population compared to UK clinical practice, when considering: 1. Disease severity markers 2. The need for initiation of empagliflozin by a specialist Therefore, the results observed in the EMPEROR-R trial may not be generalisable to UK clinical practice. **Disease severity markers** Due to the inclusion criteria, patients in the EMPEROR-R trial were more severe than would be expected in clinical practice, in particular when considering left ventricular ejection fraction (LVEF) and N-terminal pro-brain natriuretic peptide (NT-pro-BNP), two key measures of severity highlighted by the ERG (1,2). Patients in the EMPEROR-R trial had a [redacted] mean LVEF (27.2% vs. [redacted]) and [redacted] median NT-pro-BNP (1,926 pg/mL vs. [redacted] pg/mL) compared to those in the PULSE non-interventional cohort study. This is indicative of a more [redacted] patient

		population which may not be considered generalisable to those often seen in clinical practice (1). Similar results were observed when considering a subgroup of patients within the EMPEROR-R study which more closely reflects the average age to those in the PULSE study i.e., ≥65 years subgroup. Whilst these populations are more comparable with the full intention to treat (ITT) population, some key differences in the clinical prognostic indictors remain. In particular, patients in the ≥65 years subgroup in the EMPEROR-R trial had a [redacted] mean LVEF (28.0% vs. [redacted]) and [redacted] mean NT-pro-BNP (2,189 pg/mL vs. [redacted] pg/mL) compared to those in the PULSE study (3). In addition, there are key differences in the clinical history of those enrolled in the EMPEROR-R trial compared with those in PULSE. In particular, there was a [redacted] proportion of patients hospitalised for heart failure (HF) in the prior 12 months (30.7% vs. [redacted]) and patients had been diagnosed with HF for a [redacted] period (6.3 vs. [redacted])) in patients from the EMPEROR-R trial compared to those in the PULSE study. Furthermore, for the primary composite outcome and total number of hospitalisations for heart failure (HHF) (first and recurrent), there are some uncertainties in the magnitude of treatment effect with the efficacy of empagliflozin [redacted] (1). We therefore agree with the ERG that the size of the treatment benefit for empagliflozin is uncertain in the older population and the patients recruited to the EMPEROR-R trial are more severe than those seen in UK clinical practice. This suggests that there is uncertainty as to whether the results are generalisable to the UK. **The need for initiation of empagliflozin by a specialist**

		NICE NG106 outlines the process for the diagnosis and management of patients with chronic heart failure (4). Within this guideline, NT-pro-BNP levels are a key indicator used to determine the likelihood and severity of heart failure. In particular, the guideline recommends that patients with a NT-pro-BNP of between 400 and <2,000ng/L should be referred for a specialist assessment, whilst patients with a NT-pro-BNP of ≥2,000ng/L should have an urgent 2-week referral, due to very high levels of NT-pro-BNP being associated with a poor prognosis. Whilst it is common for clinical trials to enrol a population which is enriched compared to those seen in clinical practice, the patient population included within the EMPEROR-R study appears to represent a much more severe population. In particular, the median NT-pro-BNP levels seen in the EMPEROR-R trial (1,926 pg/mL in the placebo arm vs. 1,887 pg/mL in the empagliflozin arm), are close to the threshold of >2,000pg/mL that would require clinicians to make an urgent referral to a specialist within two weeks (1,4). As such, the majority of these patients would currently be reviewed by a specialist due to the concerns around the severity of their disease. In TA679 dapagliflozin is specifically recommended to be initiated on the advice of a heart failure specialist. Patients in DAPA-HF had a median NT-pro-BNP of 1,446 pg/mL in the placebo arm and 1,428 pg/mL in the dapagliflozin arm; this is lower than the patient population enrolled in the EMPEROR-R trial which is likely a more severe population. As the efficacy of empagliflozin in patients with lower NT-pro-BNP is less certain, and to ensure consistency in the NICE decision-making process, the recommendation for empagliflozin for the treatment of heart failure with reduce ejection fraction (HFrEF) should include a requirement for this treatment to be only initiated upon specialist advice, or by a heart failure specialist (5,6). We therefore agree with the ERG's comment that there is uncertainty in the generalisability of the EMPEROR-R trial and that it is a more severe patient population, the majority of which would need to be initiated in a specialist setting.

<table>
<tr><td>Key issue 2:
Uncertainty around the difference in efficacy of empagliflozin compared with standard of care in the Europe subgroup of EMPEROR-Reduced</td><td>Not applicable</td><td>No comment</td></tr>
<tr><td>Key issue 3:
Uncertainty around the efficacy of empagliflozin compared with dapagliflozin</td><td>NO</td><td>We agree that there is significant uncertainty surrounding the efficacy of empagliflozin compared with dapagliflozin.

When comparing the EMPEROR-R and DAPA-HF trials, the evidence suggests that dapagliflozin reduces the risk of cardiovascular (CV) death and all-cause mortality (ACM), whereas there is no evidence to suggest that empagliflozin has this effect (Table 1).

Specifically, the hazard ratios for time to adjudicated CV death and time to ACM in the EMPEROR-R trial are closer to unity and the confidence intervals (CI) include one, compared with those in the DAPA-HF trial, which are lower, and the CIs do not include one.

Table 1: Summary of results for empagliflozin plus SoC versus dapagliflozin plus SoC (EMPEROR-Reduced vs DAPA-HF, ITT population)
<table>
<tr><th>Endpoint: relative effect measure</th><th>EMPEROR-REDUCED: empagliflozin versus placebo</th><th>DAPA-HF: dapagliflozin versus placebo</th></tr>
<tr><td>Time to adjudicated CV death: HR (95% CI)</td><td>0.92
(0.75 to 1.12)</td><td>0.82
(0.69 to 0.98)</td></tr>
<tr><td>Time to all-cause mortality: HR (95% CI)</td><td>0.92
(0.77 to 1.1)</td><td>0.83
(0.71 to 0.97)</td></tr>
</table>
</td></tr>
</table>

		<table><tr><td>Change in KCCQ total symptom score at 8 months/7.4 months: MD (SE/95% CI)</td><td>1.6 (0.7)</td><td>2.8 (0.5)</td></tr><tr><td colspan="3">Abbreviations: CI, confidence interval; CV, cardiovascular; HR, hazard ratio; KCCQ, Kansas City Cardiomyopathy Questionnaire; MD, mean difference; SE, standard error; SoC, standard of care. Source: Adapted from empagliflozin ERG report, table 31 (1) and empagliflozin company submission, table 23 (2)</td></tr></table> Though we recognise there are differences between the EMPEROR-R and DAPA-HF trial populations, it is not appropriate to disregard differential effects on key endpoints from the trials and assume equivalent efficacy. If NICE is to make an evidence-based decision on empagliflozin, it is important that this evidence of differences in the efficacy profiles of the two sodium-glucose co-transporter 2 (SGLT2) inhibitors is considered and incorporated into an indirect treatment comparison (ITC). Whilst the company has undertaken and presented a Bucher ITC, a cost-effectiveness evaluation based upon these results has not currently been conducted. We agree with the ERG that the company should use the clinical effectiveness estimates from the Bucher ITC in the economic model to inform the treatment effectiveness versus dapagliflozin and generate an incremental cost-effectiveness ratio (ICER) for empagliflozin versus dapagliflozin. Until these steps are taken, we agree that there is significant uncertainty surrounding the efficacy of empagliflozin compared with dapagliflozin, resulting in significant uncertainty in the clinical effectiveness and cost-effectiveness estimates for this comparison.
Key issue 4:	Not applicable	No comment

The modelling of patients' distribution across the KCCQ-CSS health states		
Key issue 5: Use of a Poisson model to estimate hospitalisation for heart failure	Not applicable	No comment
Key issue 6: Overestimation of hospitalisation for heart failure in the UK population	Not applicable	No comment
Key issue 7: Modelling of mortality	Not applicable	No comment
Key issue 8: Overestimation of mortality in the UK population	Not applicable	No comment
Key issue 9: Impact of hospitalisation for heart failure in patients' quality of life	**NO**	We agree with the ERG that the impact of HHF on patients' quality of life is overestimated in the empagliflozin model. The company have implemented an annual disutility of [redacted] for each patient that experiences a HHF in the model. However, it unlikely that the reduction in a patient's quality of life after a HHF will last such a long time. Therefore, it is inappropriate to use an annual disutility and given the quality-adjusted life year (QALY) gains from HHF are a key driver in the model, this would suggest the ICER is underestimated, in favour of empagliflozin. The undiscounted utility decrement loss over a lifetime horizon is [redacted] (discounted: [redacted]) in the empagliflozin arm of the model and [redacted] (discounted: [redacted]) in the standard of care (SoC) arm (cell EP9 and FB9 of both 'ModelEngine_Empagliflozin + SoC' and 'ModelEngine_SoC' worksheets). We recommend that the impact of applying a smaller disutility in the empagliflozin model is considered and the impact of this change on the ICER.

Key issue 10: Quality of life regressions for the UK population	Not applicable	No comment
Key issue 11: Sex distribution underlying utility estimates	Not applicable	No comment
Key issue 12: Quality of life gains in EMPEROR-Reduced	Not applicable	No comment

Additional issues

Please use the table below to respond to additional issues in the ERG report that have not been identified as key issues. Please do **not** use this table to repeat issues or comments that have been raised at an earlier point in this appraisal (e.g. at the clarification stage).

Issue from the ERG report	Relevant section(s) and/or page(s)	Does this response contain new evidence, data or analyses?	Response

Additional issue 1: **Dapagliflozin is a relevant comparator for empagliflozin**	ERG report, Section 2.3	YES	We agree with the assertion in the ERG report that dapagliflozin is a relevant comparator in this appraisal and therefore should be included as a comparator in the empagliflozin economic model. The NICE Guide to the methods of technology appraisal states that "*The Committee's overall decision on whether it is a valid comparator will be guided by whether it is recommended in other extant NICE guidance, and/or whether its use is so embedded in clinical practice that its use will continue unless and until it is replaced by a new technology"* (Section 6.2.3, Page 62) (9). Dapagliflozin was recommended by NICE in February 2021 and so meets the first of these two criteria (1,5). As regards the extent to which dapagliflozin is embedded in clinical practice in the UK, SGLT2 inhibitors are recommended within the European Society of Cardiology (ESC) guidelines alongside angiotensin-converting enzyme inhibitors (ACE-I)/angiotensin-receptor blockers (ARB) and angiotensin receptor-neprilysin inhibitors (ARNI) as the foundational treatment in heart failure (10). As such, SGLT2 inhibitors are regarded as a step change in the treatment of patients with HFrEF within a highly influential international guideline. Dapagliflozin is also explicitly recommended in the 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic HF which states "*dapagliflozin or empagliflozin are recommended for patients with HFrEF to reduce the risk of HF hospitalisation and death*" (Section 2.1, Page 3609) (10). This is based on the highest class of recommendation (1), that is evidence that a given treatment or procedure is beneficial, useful and effective, and with the highest level of evidence (A), meaning data has been derived from multiple randomised clinical trials or meta-analyses (10). As well as clinical opinion and guideline recommendations, prescribing data also shows that dapagliflozin is embedded in clinical practice. We agree with the clinician expert opinion received by the ERG, that uptake of dapagliflozin is increasing and has similar levels of uptake to sacubitril/valsartan, which is considered a comparator in this appraisal. Overall, dapagliflozin meets the criteria for a comparator and should be compared against based on clinical and cost effectiveness.

			As dapagliflozin has only recently become available as a treatment option in heart failure and most prevalent patients are unlikely to have their medicine regimen changed, the increasing uptake within incident patients (i.e., newly diagnosed) should be considered. As such, prescribing patterns in newly diagnosed patients, rather than the existing prevalent population, is a better indicator as to whether dapagliflozin is embedded in UK clinical practice. We have therefore provided hospital prescribing and primary care data in this response (Figure 1, Figure 2, Figure 3 and Table 1) to aid the Committee with determining the relevance of dapagliflozin as a comparator in this appraisal. These data show: 1. **Hospital prescribing for dapagliflozin has [redacted] since the positive recommendation published by NICE in February 2021**, and at a much [redacted] rate than sacubitril/valsartan, [redacted] the hospital dispensed volume of sacubitril/valsartan in July 2021 (Figure 1). 2. **A [redacted] increase in monthly hospital prescribing of dapagliflozin by cardiologists** between February 2021 and August 2021 (Figure 2). **A [redacted] increase in monthly hospital prescribing of dapagliflozin by cardiologists** between November 2020 and August 2021 (Figure 2). 3. **An unprecedented uptake compared with sacubitril/valsartan in the primary care setting.** a. Comparing the rate of uptake of dapagliflozin with sacubitril/valsartan (Figure 3), shows that the level of growth seen for dapagliflozin since launching 10 months ago in heart failure has reached the same level that sacubitril/valsartan reached [redacted] after initiation. This indicates that the level of uptake is [redacted] across all HF patients. b. The pattern of uptake is even more pronounced if we look at data solely on initiations of dapagliflozin in primary care (Table 1). These

			data show a rapid [redacted] of dapagliflozin, [redacted] by [redacted] between February 2021 and July 2021 from 2,497 initiations to 5,824 and by [redacted] between November 2020 and July 2021. By comparison, between February 2021 and July 2021, initiations of sacubitril/valsartan in primary care increased by 40%. Moreover, initiations of dapagliflozin in primary care have now [redacted] that of sacubitril/valsartan since [redacted] (Table 2**Error! Reference source not found.**). In July 2021, [redacted] patients were initiated with dapagliflozin compared to 4,029 patients with sacubitril/valsartan. Therefore, while this dataset only represents primary care initiations, coupled with the hospital data, it provides compelling evidence that dapagliflozin is a standard of care treatment for HFrEF. Based on the evidence above, dapagliflozin is considered standard care in England and therefore is most certainly a key comparator in this appraisal. As noted in Key Issue 3, we would stress that an ICER for empagliflozin versus dapagliflozin should be generated to allow the committee to make an evidence-based decision based on the current treatment landscape. **Figure 1: Hospital dispensed volume of sacubitrl/valsartan and dapagliflozin**

			 Source: IQVIA HPAI **Figure 2: Dapagliflozin national hospital dispensed volume by key department**

			 Source: IQVIA HPA national data by department **Figure 3: Primary Care Usage of sacubitril/valsartan and dapagliflozin**

Source: IQVIA BPI (sacubitril/valsartan) and IQVIA LPD (dapagliflozin)

Table 2: New initiations of dapagliflozin and sacubitril/valsartan from primary care database

Date	Sacubitril/valsartan	Dapagliflozin	Total
Aug-20	2,126	[redacted]	[redacted]
Sep-20	2,528	[redacted]	[redacted]
Oct-20	2,772	[redacted]	[redacted]
Nov-20	2,691	[redacted]	[redacted]
Dec-20	2,802	[redacted]	[redacted]
Jan-21	2,771	[redacted]	[redacted]
Feb-21	2,882	[redacted]	[redacted]

			Mar-21 | 3,798 | [redacted] | [redacted] Apr-21 | 3,643 | [redacted] | [redacted] May-21 | 4,015 | [redacted] | [redacted] Jun-21 | 3,951 | [redacted] | [redacted] Jul-21 | 4,029 | [redacted] | [redacted] **Grand Total** | 38,008 | [redacted] | [redacted] Source: IQVIA DRx PC New Rx
Additional issue 2: **Diabetes management costs may not be completely included**	ERG report, Section 4.1.9	No	The ERG highlights that the cost of CV death in the empagliflozin model includes inpatient resource use for type 2 diabetes complications. However, it is unclear whether other costs unrelated to CV death, but associated with the management of type 2 patients, are included in the empagliflozin model. We would suggest that the company make it clear which costs are associated with the management of type 2 diabetes and which are incurred because of heart failure. The exact impact this will have on the final ICER is unknown. However, it is expected that including the costs associated with the management of type 2 diabetes will increase the final ICER, as costs will be higher in the empagliflozin arm of the model due to patients being kept alive for longer.

Summary of changes to the company's cost-effectiveness estimate(s)

Company: If you have made changes to the company's preferred cost-effectiveness estimate(s) in response to technical engagement, please complete the table below to summarise these changes.

Key issue(s) in the ERG report that the change relates to	Company's base case before technical engagement	Change(s) made in response to technical engagement	Impact on the company's base-case ICER

Not applicable	Not applicable	Not applicable	Not applicable

References

1. BMJ Technology Assessment Group (BMJ TAG). ERG Report: empagliflozin for treating chronic heart failure with reduced ejection fraction. 2021.

2. Boehringer Ingelheim. NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE Single technology appraisal Empagliflozin for treating chronic heart failure with reduced ejection fraction [ID3826] Document B Company evidence submission. 2021.

3. Boehringer Ingelheim. NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE Single technology appraisal Empagliflozin for treating chronic heart failure with reduced ejection fraction [ID3826] Clarification questions. 2021.

4. Overview | Chronic heart failure in adults: diagnosis and management | Guidance | NICE [Internet]. [cited 2021 Oct 14]. Available from: https://www.nice.org.uk/guidance/ng106

5. Overview | Dapagliflozin for treating chronic heart failure with reduced ejection fraction | Guidance | NICE [Internet]. [cited 2021 Oct 14]. Available from: https://www.nice.org.uk/guidance/ta679

6. Overview | Sacubitril valsartan for treating symptomatic chronic heart failure with reduced ejection fraction | Guidance | NICE [Internet]. [cited 2021 Oct 14]. Available from: https://www.nice.org.uk/guidance/ta388

7. Phillipo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. NICE DSU TECHNICAL SUPPORT DOCUMENT 18: METHODS FOR POPULATION-ADJUSTED INDIRECT COMPARISONS IN SUBMISSIONS TO NICE [Internet]. 2016. Available from: http://nicedsu.org.uk/wp-content/uploads/2017/05/Population-adjustment-TSD-FINAL.pdf

8. Gl DT, M U, Hs W, B P, M H, C B, et al. Health State Utilities of Patients with Heart Failure: A Systematic Literature Review. PharmacoEconomics [Internet]. 2021 Feb [cited 2021 Oct 19];39(2). Available from: https://pubmed.ncbi.nlm.nih.gov/33251572/

9. NICE. Guide to the methods of technology appraisal 2013 [Internet]. 2013. Available from: https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781

10. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC. European Heart Journal. 2021 Sep 21;42(36):3599–726.

11. Alva ML, Gray A, Mihaylova B, Leal J, Holman RR. The impact of diabetes-related complications on healthcare costs: new results from the UKPDS (UKPDS 84). Diabet Med. 2015 Apr;32(4):459–66.

Empagliflozin for treating chronic heart failure with reduced ejection fraction

ERG review of company's response to technical engagement report

November 2021

Source of funding

This report was commissioned by the NIHR Evidence Synthesis Programme as project number 13/48/48T.

1 Introduction

This document provides the Evidence Review Group's (ERG's) critique of the company's response to the technical engagement report produced by the National Institute for Health and Care Excellence (NICE) for the appraisal of empagliflozin (Jardiance®, Boehringer Ingelheim) for treating chronic heart failure with reduced ejection fraction [ID3826]. Each of the issues outlined in the technical report are discussed in further detail in Section 3.

The company's updated base case analyses are outlined in Section 2 while the ERG's analyses are reported in Section 4.

2 Updated company base case analyses

The company's updated incremental cost-effectiveness results post TE are reported in Table 1, for the trial population and in **Error! Reference source not found.** for the UK population. The company has not provided probabilistic results after TE.

Table 1. Company's deterministic base case results (discounted except for life years gained) for trial population

Interventions	Total Costs (£)	Total LYG undiscounted	Total QALYs	Incremental costs (£)	Incremental LYG undiscounted	Incremental QALYs	ICER (£/QALY)
Standard of care	£16,911	6.75	3.56	-	-	-	-
Empagliflozin	£17,633	6.89	3.70	£722	0.13	0.14	£4,999

Abbreviations: Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Table 2. Company's deterministic base case results (discounted except for life years gained) for UK population

Interventions	Total Costs (£)	Total LYG undiscounted	Total QALYs	Incremental costs (£)	Incremental LYG undiscounted	Incremental QALYs	ICER (£/QALY)
Standard of care	£15,198	6.20	3.36	-	-	-	-
Empagliflozin	£16,168	6.34	3.49	£971	0.15	0.13	£7,270

Abbreviations: Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

3 ERG review of issues

3.1 Issue 1: Uncertainty around the generalisability of the results from EMPEROR-Reduced to the older heart failure with reduced ejection fraction population expected in clinical practice

As detailed in the Evidence review group (ERG) report, the ERG and its clinical experts consider the population in EMPEROR-Reduced (hereafter referred to as EMPEROR-R) comprised of patients with more severe heart failure with reduced ejection fraction (HFrEF) than would be expected in UK clinical practice as a result of the inclusion criteria. In addition, the trial population had a mean age of ~67 years, which the ERG's clinical experts reported was approximately 10 years younger than the patients they would expect in clinical practice. The ERG noted from the age subgroup analyses in EMPEROR-R that there may be [redacted] In particular, the ERG noted that the benefit with empagliflozin compared to placebo in [redacted]. Nevertheless, the ERG considers it important to highlight that EMPEROR-R wasn't powered to detect differences in treatment effectiveness for the age subgroups and that the results of the subgroups should be interpreted with caution.

The company provided baseline characteristics for the ≥65 years age subgroup from EMPEROR-R as part of their response to technical engagement (TE) and the ERG considers the ≥65 years age subgroup shows some differences at baseline compared to the full intention-to-treat (ITT) population (Table 1). The ERG notes that the baseline age in the EMPEROR-R ≥65 years age subgroup is [redacted] The ERG also notes that the EMPEROR-R ≥65 years age subgroup has a slightly higher N-terminal pro hormone B-type natriuretic peptide (NT-proBNP) and patients have slightly worse baseline renal function in terms of the proportion with baseline estimated glomerular filtration rate (eGFR) ≥ 60 ml/min/1.73m^2 compared to the full ITT population. The ERG thus considers the ≥65 years age subgroup from EMPEROR-R still comprise of a more severe HFrEF population than that expected to be eligible for empagliflozin in United Kingdom (UK) clinical practice and [redacted] However,

given the closer match in age to the UK population, the ERG considers the ≥65 years age subgroup from EMPEROR-R to be a subgroup of importance.

The ERG acknowledges the company argument that there is a large amount of missing data from PULSE and the reasons for missing data are not known but the ERG also notes that PULSE comprises a [redacted] UK patients with HFrEF. The ERG also considers it important to flag that, as noted in the ERG report, empagliflozin is recommended for treatment of patients with symptomatic HFrEF and thus EMPEROR-R does not include asymptomatic (New York heart association [NYHA] Class 1) patients, [redacted] The ERG's clinical experts reported that up to 30% of HFrEF patients may be asymptomatic (NYHA Class 1) but unfortunately PULSE did not contain complete data on NYHA classification and so it is not clear what proportion of patients in PULSE would potentially be ineligible for empagliflozin. However, given the likely small proportion of patients who are asymptomatic and the absence of other data, the ERG considers the use of the HFrEF subgroup data from PULSE to be reasonable.

The ERG also notes that the company argues that the HHF events in EMPEROR-R were independently adjudicated, whereas in PULSE this was not the case, and so it is possible that there may be some errors in coding of hospitalisation events for some patients with HHF incorrectly recorded for non-HHF events and vice versa. The extent of these potential coding errors in PULSE is unknown and therefore it is not possible to predict the direction of any resulting bias. Nevertheless, the ERG considers that given the data from PULSE were all collected in the UK it is likely that broadly standard definitions have been used to classify HHF events. The ERG therefore considers PULSE remains the best available source of evidence to validate heart failure (HF) outcomes in the UK population.

Table 3. Baseline characteristics of EMPEROR-R (ITT and 65+ years) vs PULSE (Adapted from company response to technical engagement, Table 1).

Treatment Arm	EMPEROR-R (ITT; Combined Groups)	EMPEROR-R (65+; Combined Groups)	PULSE[a]
N	3,730	2315	[redacted]
Age [mean (SD)]	66.8 (11.0)	73.8 (5.9)	[redacted]
Age at baseline >= 65 years [% (N)]	62.1% (2315)	100.0% (2315)	NR

Male	76.1% (2837)	76.1% (1762)	[redacted]
Race [% (N)]			
White	70.5% (2629)	75.6% (1750)	[redacted]
Asian	18.0% (672)	15.9% (369)	[redacted]
Native	1.0% (39)	0.9% (21)	NR
Black	6.9% (257)	4.6% (106)	[redacted]
Multiple/mixed	1.6% (61)	1.0% (24)	[redacted]
Pacific	0.4% (14)	0.3% (8)	NR
Region [% (N)], global			
Asian	13.2% (493)	13.0% (301)	NR
Europe	36.3% (1353)	42.7% (988)	[redacted]
Latin America	34.5% (1286)	29.2% (677)	NR
North America	11.4% (425)	12.2% (282)	NR
Other	4.6% (173)	2.9% (67)	NR
Type-2 Diabetes [% (N)]	49.8% (1856)	49.0% (1134)	[redacted]
BMI, kg/m^2 [mean (SD)]	27.9 (5.4)	27.5 (5.1)	[redacted]
BMI, kg/m^2 [% (N)] missing	NR	NR	[redacted]
eGFR at baseline >= 60 ml/min/1.73m2 [% (N)]	51.8% (1931)	40.1% (929)	[redacted]
eGFR at index, [% (N)] missing	NR	NR	[redacted]
Prior hospitalisation for HF in 12 months prior [% (N)]	30.9% (1151)	29.2% (675)	[redacted]
Prior atrial fibrillation or flutter [% (N)]	38.6% (1441)	46.5% (1076)	[redacted]
Time since diagnosis (in years)			
Mean, years (SD)	6.14 (6.32)	6.69 (6.62)	[redacted]
0–1	18.6% (692)	15.8% (366)	NR
1–5	37.9% (1415)	36.6% (848)	NR

5+	43.5% (1623)	47.6% (1101)	NR
Ischemic cause of HF [% (N)]	51.7% (1929)	57.1% (1322)	[redacted]
LVEF [mean (%)]	26.8	30.2	[redacted]
LVEF, [% (N)] missing	NR	NR	[redacted]
Heart rate (bpm) [mean (SD)]	71.3 (11.7)	70.3 (11.5)	[redacted]
Heart rate (bpm), [% (N)] missing	NR	NR	[redacted]
NT-proBNP, pg/mL, [mean (SD)]	3034.7 (3665.5)	3226.9 (3788.5)	[redacted]
NT-proBNP, pg/mL, median (IQR)	1910 (1115-3481)	2047 (SD)	[redacted]
NT-proBNP, pg/mL, [% (N)] missing	NR	NR	[redacted]
Blood pressure (mm Hg) [mean (SD)]	122.0 (15.6)	123.2 (15.7)	[redacted]
Blood pressure (mm Hg), [% (N)] missing	NR	NR	[redacted]
Implantable cardioverter-defibrillator [% (N)]	22.8% (851)	23.5% (543)	[redacted]
Cardiac resynchronisation therapy [% (N)]	11.9% (442)	15.0% (347)	NR
HF medication [% (N)]			
ACEI or ARB + BB (no IVA, no ARNI)	62.7% (2339)	62.9% (1456)	[redacted]
ARNI + BB (no IVA, no ACEI or ARB)	15.6% (580)	15.6% (360)	
Other treatment	21.7% (811)	21.6% (499)	[redacted]
EQ-5D score [mean (SD)]	0.7 (0.2)	0.7 (0.2)	NR

Notes: a. As noted in Company Submission Appendix O, the baseline characteristics are for the HFrEF-prevalent and incident population at index (2015).

b. In PULSE, ischemic heart disease included myocardial infarction, coronary procedure and other IHD.

c. Primary care prescribing data only for PULSE – the company considers this to be a substantial underestimation.

Abbreviations: ARB, angiotensin receptor blocker; ACEi; angiotensin converting enzyme inhibitors; ARNI, angiotensin receptor neprilysin receptors; BMI, body mass index; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; EQ-5D, EuroQol- 5 dimension; HF, heart failure; IHD, ischemic heart disease; IQR, interquartile range; ITT, intention-to-treat; IVA, ivabradine; LVEF, left ventricular ejection fraction; NR, not reported; SBP, systolic blood pressure; SD, standard deviation.

Please see Issue 6 for more detail on the ERG view of the company's modelling of HHF and Issue 8 for the modelling of mortality.

3.2 Issue 2: Uncertainty around the difference in efficacy of empagliflozin compared with standard of care in the Europe subgroup of EMPEROR-Reduced

In response to technical engagement the company provided subgroup baseline characteristics for the Europe region subgroup from EMPEROR-R as well as results for the composite renal outcome, all-cause mortality and Kansas City Cardiomyopathy Questionnaire clinical summary score (KCCQ-CSS). The ERG discussed the baseline characteristics for the Europe subgroup in the ERG report Section 3.3.10.2 and noted there were differences between the ITT population and Europe subgroups including a slightly higher mean age and that the subgroup appeared to comprise of a more severe population compared to the ITT population.

The results for the Europe region subgroup outcome of all-cause mortality presented in the company's response to technical engagement are [redacted]

For the composite renal outcome, the Europe subgroup results [redacted]

As detailed in the ERG report, the results of the Europe subgroup for the primary composite outcome, total HHF and cardiovascular (CV) mortality [redacted] with empagliflozin compared to the overall EMPEROR-R ITT population (ERG report, Table 26). The ERG considers it important to highlight that the study was not powered to detect statistically significant differences in subgroups and thus caution should be taken in drawing conclusions from any of the Europe region subgroup results. Nevertheless, the ERG considers it important to highlight that there may be a difference in efficacy with empagliflozin related to geographic region.

Table 4. Composite renal outcome and all-cause mortality results for EMPEROR-R ITT and Europe subgroup (adapted from Company response to technical engagement, Table 3 and Table 4)

Clinical Endpoint	EMPEROR-ITT Empagliflozin	EMPEROR-ITT placebo	EMPEROR-R Europe Empagliflozin	EMPEROR-R Europe placebo
Composite renal outcome[a]				
Patients with the composite renal event, N (%)	31 (1.6)	58 (3.1)	[redacted]	[redacted]
Incidence rate per 100 patient years at risk	1.56	3.07	[redacted]	[redacted]
Hazard ratio vs. placebo (95% confidence interval), composite renal outcome	0.50 (0.32 - 0.77)		[redacted]	
Nominal p-value	0.0019		[redacted]	
All-cause mortality				
Number of patients with the event, N (%)	249 (13.4)	266 (14.2)	[redacted]	[redacted]
Incidence rate [patients with events per 100 patient years at risk]	10.17.6	10.78.1	[redacted]	[redacted]
Hazard ratio vs. placebo (95% CI), all-cause mortality	0.92 (0.77-1.10)		[redacted]	
Nominal p-value	0.35		[redacted]	
[redacted]				

Table 5. [redacted]

Mean change from Baseline to:	ITT population			Europe subgroup		
	Empagliflozin (10mg)	Placebo	Difference	Empagliflozin (10mg)	Placebo	Difference
Baseline						
N	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mean (SE)	[redacted]	[redacted]		[redacted]	[redacted]	
Week 12[a]						
Adjusted Mean, SE, [95% CI], P [a]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 32 [a]						
Adjusted Mean, SE, [95% CI], P [a]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 52 [a]						
Adjusted Mean, SE, [95% CI], P [a]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]						

Abbreviations: ITT, intention-to-treat; SE, standard error.

3.3 Issue 3: Uncertainty around the efficacy of empagliflozin compared with dapagliflozin

As detailed in the ERG report, the ERG considers the results of the Bucher ITC to be [redacted] and that the company's assumption of [redacted] for empagliflozin and dapagliflozin lacks robustness. The ERG is particularly concerned that the company has made an assumption of [redacted] based on the results of a Bucher indirect treatment comparison (ITC) and a pooled meta-analysis comprising of only a single trial for each intervention (EMPEROR-R for empagliflozin and DAPA-HF for dapagliflozin). The results from the company's Bucher ITC showed [redacted] (Company submission Table 23 and Company response to technical engagement, Table 6). [redacted] as noted in the ERG report, the ERG considers the results of the Bucher ITC show [redacted] with empagliflozin compared to dapagliflozin (HR [empagliflozin vs dapagliflozin] [redacted]; 95% CI: [redacted], and HR [redacted]; 95% CI: [redacted], respectively) and a [redacted] mean change in Kansas City Cardiomyopathy Questionnaire total symptom score (KCCQ-TSS) score from baseline to 8 months with dapagliflozin (mean difference [MD] 2.8) compared with empagliflozin (MD 1.6). In addition, the Bucher ITC of worsening renal function using the DAPA-HF definition suggested that empagliflozin might be [redacted] in reducing the hazard of worsening renal function compared to dapagliflozin, although the [redacted] (HR [redacted]; 95% CI: [redacted]).

In the company response to technical engagement a figure was provided showing the naïve comparison between empagliflozin and dapagliflozin for CV mortality using the Kaplan-Meier (KM) data from EMPEROR-R and DAPA-HF (Figure 1). The ERG considers the placebo data in the two trials [redacted] between empagliflozin and dapagliflozin, [redacted] However, the ERG also considers it important to highlight the high risk of bias likely to exist within a naïve comparison and thus considers the ITC a more robust method of comparison between empagliflozin and dapagliflozin.

Figure 1. [redacted]

[redacted]

The ERG notes that KCCQ-CSS is used in the economic model rather than KCCQ-TSS and so the ERG explored conducting a Bucher ITC for KCCQ-CSS. The ERG was only able to replicate the company's analysis for KCCQ-TSS when assuming the reported SEs are standard deviations (SDs). As such, the ERG is concerned that the company may be underestimating the uncertainty in the analysis of KCCQ-TSS, although the mean estimate is unchanged.

In the ERG's analysis of KCCQ-CSS the results were [redacted]

The results from the company's Bucher ITCs were not used in the economic model in the company submission and instead the company [redacted] Therefore, the NICE technical team and the ERG requested the results of the

Bucher ITC to be included in the economic model to undertake a cost-effectiveness analysis of empagliflozin vs dapagliflozin.

The company noted the ERG's view that there is no evidence to substantiate the company's original assumption of a constant treatment effect (and therefore the use of a HR) for empagliflozin over standard of care (SoC) on CV survival throughout the model and concluded the same could be argued for empagliflozin vs dapagliflozin. The company added that a naïve comparison of the empagliflozin and dapagliflozin KM CV survival curves showed that they hardly separated during the follow-up period of EMPEROR-R and DAPA-HF and that the HR estimated in the Bucher ITT was [redacted]

The ERG acknowledges the company's observation that the HR for CV survival ([redacted] However, the ERG notes that the [redacted] As discussed above, given that the Bucher ITC only included two studies, it would be expected that [redacted] Therefore, the ERG's preference is to include the Bucher HRs on mortality for empagliflozin vs dapagliflozin although the ERG appreciates that [redacted] Assuming the [redacted] results are evidence of equivalence results in the two treatments having identical costs and benefits. However, the ERG has supplied an analysis assuming that the mean estimates accurately reflect the differences between treatments to aid committee in their decision making.

The company also did not use the Bucher HR for mean KCCQ-TSS of [redacted] which indicates that dapagliflozin might be [redacted] than empagliflozin at preserving patents' KCCQ-TSS, albeit the ERG does not consider that the company has appropriately captured the magnitude of the uncertainty (as discussed above). The results of the ERG's analysis of KCCQ-CSS results were [redacted]

The company's model structure means that it is not possible to translate differences in mean KCCQ-CSS into state transitions in the model. Therefore, the ERG had no way of implementing the mean KCCQ-CSS [redacted] indicated by the Bucher analysis in the model. Thus, the ERG notes that the

company's assumption that the impact of empagliflozin and dapagliflozin on preserving KCCQ-CSS is the same is a possible [redacted] of the dapagliflozin effect on preserving KCCQ-CSS.

The company's analysis included a [redacted] associated with empagliflozin for HHF and renal outcomes. The Bucher HRs used in the model for the empagliflozin vs dapagliflozin were of [redacted] for HHF and [redacted]) for renal outcomes. The ERG believes that the company has incorrectly used the [redacted] HR and instead meant to use the [redacted] HR reported in Table 6 of the company's response to TE. Therefore, the ERG replaced this in the model.

The company reports a dominant base case incremental cost-effectiveness ratio (ICER) in favour of empagliflozin, when no survival benefit is assumed between empagliflozin and dapagliflozin. The company's results (corrected by the ERG) show savings of £166 per patient; largely driven by improvements in renal function and no difference in quality adjusted life years [QALYs] – **Error! Not a valid bookmark self-reference.**). The company's PSA ICER was also dominant in favour of empagliflozin.

The ERG conducted additional analyses to show the impact of adding the different HRs from the Bucher ITC on the ICER for empagliflozin vs dapagliflozin. The results in **Error! Not a valid bookmark self-reference.** for the ITT population show that the inclusion of the survival benefit associated with dapagliflozin is by far the biggest QALY driver of the analysis, with HHF and renal outcomes only very marginally affecting the incremental QALYs. The key drivers of costs are the inclusion of the survival benefit for dapagliflozin as well as the inclusion of the benefit associated with empagliflozin on renal outcomes. The inclusion of HHF outcomes has a negligible impact on the final ICER. The ERG's preferred ICER for empagliflozin vs dapagliflozin in the ITT population amounts to £5,910 (with empagliflozin being less costly and less effective than dapagliflozin).

Table 7 presents the results for the >65-year population. The conclusions are the same as those inferred for the ITT population, with the ERG's preferred ICER amounting to £5,657 (empagliflozin less costly and less effective than dapagliflozin).

When the ERG ran the PSA including a survival benefit for dapagliflozin, the resulting ICER was £4,069 with empagliflozin being less costly (-£563) and less effective than dapagliflozin (-0.09 QALYs). The ERG's results in Figure 2 demonstrate that 64% of iterations were in the south-western quadrant (empagliflozin is less costly and less effective than dapagliflozin), while 18% of the PSA iterations produced an ICER in the north-eastern quadrant of the cost-effectiveness plane.

When the ERG ran the PSA including a survival benefit for dapagliflozin for the >65-year population, the resulting ICER was £5,908 with empagliflozin being less costly (-£584) and less effective than dapagliflozin (-0.10 QALYs). The ERG's results in Figure 3 demonstrate that 66% of iterations were in the south-western quadrant (empagliflozin is less costly and less effective than dapagliflozin), while 18% of the PSA iterations produced an ICER in the north-eastern quadrant of the cost-effectiveness plane.

Table 6. Incremental changes to ICER for empagliflozin vs dapagliflozin ITT population (with ERG's correction)

	Results per patient	Empagliflozin	Dapagliflozin	Incremental value
	Company's base case corrected (assuming HHF benefit for empagliflozin + assuming renal benefit for dapagliflozin)			
	Total costs	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)	[redacted]	[redacted]	[redacted]
0	Assuming equal effectiveness between treatments in all outcomes			
	Total costs	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)	[redacted]	[redacted]	[redacted]
1	Assuming survival benefit for dapagliflozin			
	Total costs	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]

	ICER (£/QALY)			[redacted]
1+2	Assuming survival benefit for dapagliflozin + Assuming HHF benefit for empagliflozin			
	Total costs	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			[redacted]
1+2+3	Assuming survival benefit for dapagliflozin + assuming HHF benefit for empagliflozin + Assuming renal benefit for empagliflozin			
	Total costs	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			[redacted]

Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

*[redacted]

Figure 2. Probabilistic scatter plot of ICERs for empagliflozin vs dapagliflozin (ITT population)

[redacted]

Table 7. Incremental changes to ICER for empagliflozin vs dapagliflozin >65 years population (with ERG's correction)

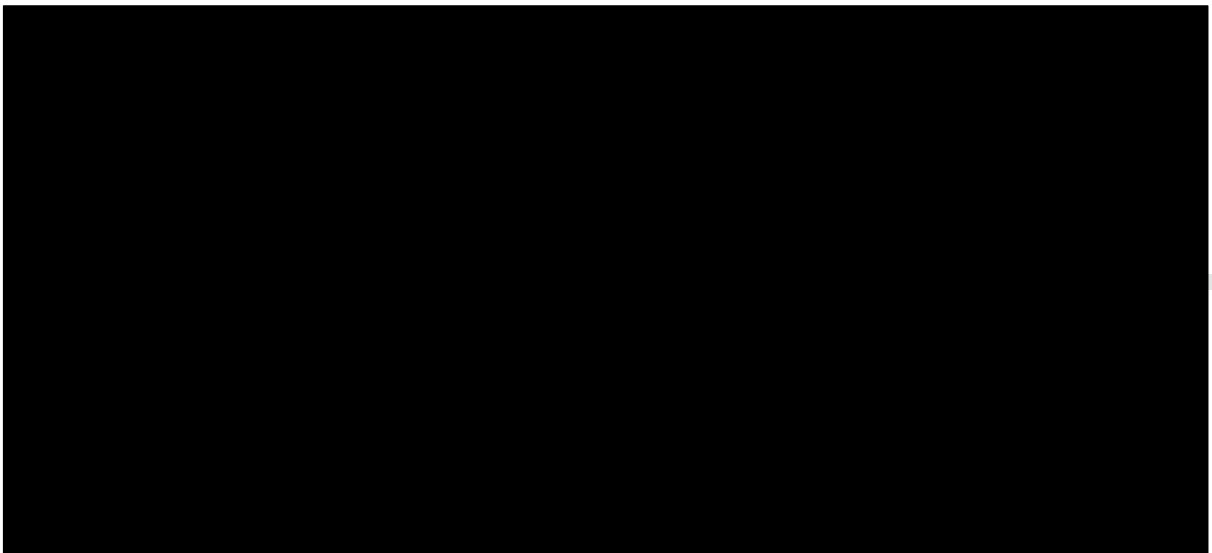
	Results per patient	Empagliflozin	Dapagliflozin	Incremental value
	Company's base case corrected (assuming HHF benefit for empagliflozin + assuming renal benefit for dapagliflozin)			

	Total costs	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)	[redacted]	[redacted]	[redacted]
0	Assuming equal effectiveness between treatments in all outcomes			
	Total costs	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)	[redacted]	[redacted]	[redacted]
1	Assuming survival benefit for dapagliflozin			
	Total costs	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			[redacted]
1+2	Assuming survival benefit for dapagliflozin + Assuming HHF benefit for empagliflozin			
	Total costs	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			[redacted]
1+2+3	Assuming survival benefit for dapagliflozin + assuming HHF benefit for empagliflozin + Assuming renal benefit for empagliflozin			
	Total costs	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			[redacted]

Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

*South-western quadrant ICERs - Less costly and less effective

Figure 3. Probabilistic scatter plot of ICERs for empagliflozin vs dapagliflozin (>65 years population)



3.4 Issue 4: The modelling of patients' distribution across the KCCQ-CSS health states

During TE the ERG asked that the company provided additional details on the following issues:

1. Which dataset from EMPEROR-R is being used to estimate the transition probabilities (TPs) across the different KCCQ-CSS states in the model – the company confirmed that the results were based on the randomised set in EMPEROR-R with observed cases, including data after treatment discontinuation, which the ERG notes is consistent with the other clinical inputs from EMPEROR-R used in the model.
2. The ERG asked that the company provided the data from EMPEROR-R that allowed the estimation of TPs and proportion of patients in each KCCQ-CSS in the model – the company provided the observed TPs from the trial (Table 10 and Table 11 of the company's response to TE) for patients' movements across KCCQ-CSS quartiles. The ERG used these data in combination with the proportion of patients in each KCCQ-CSS category at baseline in the trial to replicate the proportion of patients at week 12, 32 and 52 in the trial (provided in Table 35 in the company's clarification response). The ERG validated the values provided by the company and is more reassured that the proportion of patients moving across the KCCQ-CSS states in the model (also provided in Table 35) closely matches the proportions seen in the trial.

   The ERG asked that the company provided the TPs observed in EMPEROR-R for the KCCQ-CSS quartiles defined in the model and explained how these related to the mean changes reported in the trial – the ERG's request was to obtain the raw data from the trial to be able to relate the mean change in KCCQ-CSS scores (reported as one of the key clinical outcomes of the EMPEROR-R trial) with the clinical data used in the model. Even though the ERG agrees with the company that mean KCCQ-CSS changes cannot be

compared with proportion of patients, the ERG sought to understand the data behind both outcomes.

The company provided the values reported in Table 8, which shows the distribution of mean changes in KCCQ-CSS from baseline to week 12 in EMPEROR-R. The ERG notes that these data are helpful in understanding the connection between the mean KCCQ-CSS changes reported in trial from baseline to week 12 ([redacted] for placebo and empagliflozin, respectively) with the model inputs. The results show, for example, that 10% of patients in the placebo arm had a decrease of mean KCCQ-CSS of more than [redacted] points while 25% of patients had improved by more than [redacted] points with empagliflozin. The ERG considers that it would be helpful to the committee if the company could provide the equivalent results for the other periods of analysis from the trial (i.e., week 12 to week 32 and week 32 to 52).

Table 8. Distribution of individual changes in KCCQ-CSS from baseline to W12 visit (without imputation)

	N	Mean	SD	Min	P10	P25	Median	P75	P90	Max
Empagliflozin 10mg	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Placebo	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

3. Finally, the ERG asked that the company conducted scenario analyses where the effect of empagliflozin on KCCQ-CSS (sustained by the combination of the proportion of patients in the better KCCQ-CSS states in the empagliflozin arm at month 8 and the low probability of disease progression for both SoC and empagliflozin arms in month 9+) was waned over time in the model.

The company's response cited an empagliflozin trial by Zinman *et al.* 2015, where T2DM patients were randomised to either empagliflozin or placebo. The company reported that the reduction seen in all-cause mortality, CV-mortality and heart failure-related hospitalisation for empagliflozin patients was sustained for 3.1 years. The company added that in the dapagliflozin appraisal in HFrEF (TA679), the Committee concluded that, "*there was no evidence for or against treatment waning in the long-term. Clinical*

experts and stakeholders confirmed that treatment with dapagliflozin would likely be lifelong".

The ERG notes that there are two slightly different issues related to the long-term effect of empagliflozin – the effect of the drug after patients discontinue treatment and the long-term effect of the drug for patients staying on life-long treatment.
The ERG agrees that the probability of patients discontinuing treatment with empagliflozin is reduced and similar to that of dapagliflozin (about 16% according to the ERG's clinical experts and the EMPEROR-R trial results), however, the ERG disagrees with the company's underlying modelling assumption that patients still benefit from empagliflozin after they discontinue treatment. Given that the there is a higher percentage of empagliflozin patients in the highest KCCQ-CSS state in the model at month 8, and that the TPs used in month 9+ of the company's model assume that all patients (on treatment and off treatment) have a very small probability of leaving the KCCQ-CSS state they are in at month 8, the benefit associated with empagliflozin is broadly maintained for patients discontinuing treatment after month 8 of the model. The company's assumption that empagliflozin patients experience SoC TPs after discontinuation is only partially conservative and leads to a sustained relative treatment effect for patients in KCCQ-CSS 4 in the model over time.

For the 84% of patients not discontinuing treatment in the model, the ERG notes that the validity of the company's assumption that the changes seen in EMPEROR-R from month 8 - 13 are sustained for approximately 30 years in the model remains uncertain. Due to the company's model structure, this assumption impacts the benefits associated with empagliflozin on HHF and mortality, as these outcomes are dependent on patients' distribution across KCCQ-CSS states. The ERG notes that in the Zinman *et al.* 2015 cited by the company, the mean time on treatment for empagliflozin patients was 2.6 years while the mean follow-up was 3 years, therefore the study does not provide any additional evidence on the long-term impact of empagliflozin after treatment discontinuation.

As a response to the ERG request, the company undertook a scenario analysis where the proportion of patients in the KCCQ-CSS quartiles under the treatment arm was set equal

to those proportions in the placebo arm at 5, 3, 2, and 1 years. The company reported the changes in the ICERs for these scenarios and concluded that the impact was small. The ERG agrees with the company, however, notes that these scenarios are more meaningful when ran in combination with the scenarios varying the assumptions around the survival benefit associated with empagliflozin (issue 7).
The company also included an additional scenario in the model where the TPS between KCCQ-CSS quartiles for patients on treatment were assumed to be the same as those for patients off treatment from month 9 onwards in the model.

The ERG conducted a range of scenarios using different permutations including these 3 additional scenarios and presents the results in Section 4.

3.5 Issue 5: Use of a Poisson model to estimate hospitalisation for heart failure

Given that time to HHF KM data were available from EMPEROR-R, the ERG requested that the company used these data to model time to HHF. Using the KM HHF data from EMPEROR-R would have allowed the company to fit a parametric survival curve to the data and extrapolate into the model's time horizon without having to assume a constant rate of HHF and without having to assume a constant treatment effect with empagliflozin. The ERG also noted that using KM data for time to HHF would have allowed the company to model time to first and subsequent HHF separately and that this could be of importance given the results reported in the EMPEROR-R CSR, indicating that at 2 years, [redacted] of patients in the empagliflozin arm had experienced a second HHF, while [redacted] of patients had experienced a second event in the placebo arm.

Furthermore, using KM data for time to HHF would have allowed the company to undertake a scenario analysis where HHF KM data from the >65 subgroup in EMPEROR-R was adjusted to reflect a lower number of total HHFs in the model based on the HHF predictions from PULSE (see issue 6).

In their response to TE, the company reported using KM data on time to HHF from EMPEROR-R and re-arranging the data by "*using a counting process setup with start and stop times to create periods defined by the occurrence of each hospitalisation and/or changes in KCCQ-CSS quartiles. That is, a patient will have one record per change in KCCQ-CSS and per hospitalisation with start and stop times of the period defined by the time when these changes occur*." The company then fitted parametric models to these data with *flexsurvreg* in R testing exponential, Weibull, Gompertz, log-logistic, log-normal and generalised gamma models. The company stated that it was inappropriate to

use the parametric models fitted to the data as these showed a declining rate in hospitalisation over time (which was deemed clinically implausible) due to the declining number of patients at risk near the end of the KM curves.

The company concluded that its findings supported the use of a fitted Poisson model which assumed a constant rate of hospitalisation, while allowing the rate to increase as patients progress over time through inclusion of time-varying KCCQ-CSS as a predictor. The company added a scenario analysis in the model where a constant increase in the rate of HHF was assumed. Results from this scenario suggested that allowing the rate of HHF to increase over time reduces the ICER due to higher reduction in number of HHFs.

The ERG appreciates that the company's model was structured to use total number of HHF events, therefore, requiring a transformation of KM HHF data into count data. Nonetheless, the ERG notes that this method is less robust in translating the effect of empagliflozin on HHF observed in EMPEROR-R. The ERG used the KM data provided by the company during TE to produce KM curves for time to first HHF and second HHF for the ITT population in EMPEROR-R. Figure 4 shows that while it took longer for empagliflozin patients to experience a first hospitalisation, the opposite was true for second hospitalisation events, where fewer SoC patients had had an event at year 1 and year 2 compared to empagliflozin patients (total percentage of patients experiencing second events in EMPEROR-R was 35% for empagliflozin and 33% for SoC patients). Given the company's use of total number of events in each arm to model HHF in the model, this change in trend for time to first and time to second HHF was not captured in the economic analysis.

Figure 4. Time to HHF first and second events observed in EMPEROR-R (ITT population)

The ERG did not have time to conduct a complete fitting exercise of parametric models to the KM HHF data observed in EMPEROR-R, however, preliminary analysis showed that the underlying hazard functions in the parametric models fitted to the KM HHF data from EMPEROR-R were not monotonically decreasing across all models. This compares to the company's assessment that the fitted parametric models to the re-arranged HHF data (per occurrence of each hospitalisation and/or changes in KCCQ-CSS quartiles per interval of time) were not appropriate to use in the model given their underlying decreasing hazard rate.

The ERG maintains its view that using the KM HHF data from EMPEROR-R (independent from KCCQ-CSS states) would have allowed the company to estimate long-term HHF by relying on observed data and not assuming a constant rate of HHF. Importantly, the use of KM data would have allowed the company to model time to first and subsequent HHF separately and that this could be of importance given the results seen in Figure 4.

Finally, the ERG acknowledges that the company's approach accurately reproduces the number of HHFs observed in the trial over 18 months and also agrees with the company's assessment that increasing the rates of HHF overall in the model ultimately benefits empagliflozin as the potential benefit for reducing HHF events also increases. Nonetheless, the ERG remains uncertain if HHFs are accurately estimated in the long-term model for the trial population and notes that using the observed data extrapolated into the future might not have resulted in an overall increased rate of HHF in both arms of the model. Importantly, if first and second events were to be separated in the model, it is likely that the benefit associated with empagliflozin on HHF would reduce.

3.6 Issue 6: Overestimation of hospitalisation for heart failure in the UK population

The company conducted a scenario analysis where the subgroup data from EMPEROR-R for patients above 65 years (mean age 74 years) were used in the model to try and reflect the lower rates of HHF seen in PULSE and in clinical practice. Nonetheless, the analysis conducted by the company still grossly overestimated HHF in the model when compared to PULSE, likely due to the trial's inclusion of sicker patients.

The ERG recommended that the company undertook a scenario analysis where HHF KM data from the >65 subgroup in EMPEROR-R was used to model time to HHF in the UK population and that the respective extrapolated curves were adjusted to reflect a lower number of total HHFs in the model (based on the HHF predictions from PULSE) with, for example, the use of a HR.

In their response to TE, the company stated that the difference in outcomes (i.e., HHF and CV-mortality rates) observed in EMPEROR-R and PULSE is likely due to inaccurate recording of events in PULSE rather than a clinically meaningful difference in the patients' characteristics. As discussed in Section 3.1, the ERG reiterates its view that PULSE remains the best available source of evidence to validate HF outcomes in the UK population.

Similar to issue 5, the company reported using the re-arranged KM data on time to HHF and fitting parametric models to these data for the >65 subgroup of EMPEROR-R. The company reported that "*the use of parametric modelling for recurrent HHF events does not offer an improvement over the Poisson modelling framework and* [the latter] *is retained as the approach in the model*."

The company did not provide the raw KM data for the >65 subgroup of EMPEROR-R, therefore, the ERG could not undertake the equivalent analysis of that done for the ITT population.

In order to adjust the Poisson model to the PULSE HHF data, the company fitted a joint regression model of the individual patient-level data from the PULSE and EMPEROR-R and included a term in the model for the population source (PULSE vs EMPEROR-Placebo). This was done with a negative binomial model for the total number of HHF events observed for patients with follow-up duration as an offset (log-transformed). The company reports that a negative binomial model was used as the latter had been used in the original analyses of the PULSE data to address overdispersion in event counts.

The company's negative binomial model yielded a rate-ratio of 0.43 when comparing patients in PULSE with HFrEF and patients aged 65 years or older from the placebo arm in EMPEROR-R. The company included a scenario in the model where the HHF Poisson model was adjusted by the 0.43 ratio, resulting in a lower number of HHF for the >65 years population (rate per 100 patient-years decreased from 18.44 for SoC patients and 16.27 for empagliflozin patients to 7.93 and 7, respectively). When compared to the PULSE results, the SoC arm of the model estimated 14,208 HHF events vs the 16,033 events observed in PULSE (over the first three years in both sources). As

expected, reducing the total number of HHF in the model increased the ICER from £7,270 to £9,678 QALY gained.

The ERG acknowledges that the company's scenario analysis more accurately reproduces the number of HHFs observed in the first three years of PULSE. Nonetheless, the ERG remains uncertain if HHFs are accurately estimated in the long-term model for the UK population given the company's assumption of a constant HHF rate throughout the analysis. The ERG notes, again, that the HHF KM data for the >65 subgroup of EMPEROR-R was not provided by the company therefore, the ERG could not undertake the equivalent analysis of that done for the ITT population.

3.7 Issue 7: Modelling of mortality

The ERG considered that the company's original assumption of a constant treatment effect (and therefore proportional hazards [PHs]) of empagliflozin over SoC throughout the model was unsubstantiated, both for all-cause and CV-related death. The ERG noted that the empagliflozin and placebo KM OS curves hardly separated during the follow-up period of EMPEROR-R and that the HRs in the trial for all-cause and CV-related death were not statistically significant (and also signaled a small effect size). Consequently, the ERG conducted a scenario analysis where CV and non-CV mortality were assumed to be the same in the empagliflozin and the SoC arms of the model. The ERG noted that when no treatment effect was assumed for empagliflozin there was still a benefit associated with empagliflozin on both CV and non-CV mortality because the probability of patients dying was different in every KCCQ-CSS state of the model. Given that patients in the empagliflozin arm of the model have a higher probability of remaining in the better KQCC-CSS states over time compared with SoC patients, the former also experience a lower probability of death.

The ERG recommended that the company considered adding a scenario analysis in the model where it was assumed that empagliflozin had no survival benefit over SoC (including through the residency in KCCQ-CSS states).

As a result of TE, the company has adopted a new base case, where no direct survival benefit was assumed for empagliflozin. Nonetheless, the company noted that as in EMPEROR-R empagliflozin patients had a higher probability of remaining in better KCCQ-CSS states over time compared with SoC patients, it was reasonable to maintain this indirect survival benefit in the model.

As a scenario analysis, the company implemented an option in the model where it was assumed that empagliflozin has no direct or indirect survival benefit over SoC. A Weibull function was fitted to the

OS KM data for the placebo arm of the EMPEROR-R trial with no KCCQ-CSS predictors, for the ITT and 65+ subgroups, separately. These Weibull curves were then applied in both arms in the model to capture mortality, for the ITT and 65+ subgroups, separately.

The ERG considers that any potential impact of residency in better KCCQ-CSS states on survival would have been captured through the OS curves for empagliflozin vs SoC. The ERG reiterates its original view that the OS curves from EMPEROR-R do not provide sufficient evidence to substantiate empagliflozin having an impact on patients' survival compared to SoC patients. Therefore, the ERG's preference is to use the company's updated base case assumption of no direct survival benefit with empagliflozin, combined with the company's scenario analysis where survival is the same in both treatment arms, regardless of KCCQ-CSS residency in the model.

The ERG also notes its' clinical experts' opinion provided originally that while the Weibull curves fitted by the company were appropriate for the >65 years population, the long-term predictions of the Gompertz curves would be a better representation of the higher mortality seen in the trial population. The company has only provided the Weibull curves for both populations, therefore, the ERG could not use a Gompertz curve to estimate the impact of removing all the survival benefits associated with empagliflozin for the trial population.

Removing all the survival benefit associated with empagliflozin in the model and using a Weibull model to estimate survival in the ITT and in the >65 years population led to a change in ICERs from £4,999 to £4,777 and from £7,270 to £9,780 , respectively.

The ERG notes that for the ITT population, the ICER decreased when the survival benefit for empagliflozin was removed from the model. This result is counterintuitive and is due to the increase in costs in the SoC arm (£424) being higher than the increase in costs in the empagliflozin arm (£94), leading to a decrease of £330 in the incremental costs associated with empagliflozin when no survival benefit was assumed in the model. Overall, there was a decrease in the QALY gain associated with empagliflozin (0.08 vs 0.14) when the survival benefit was removed from the model. The ERG did not have the time to explore this further but considers it likely to be due to the change in survival curves for the ITT population led to this idiosyncrasy in results.

3.8 Issue 8: Overestimation of mortality in the UK population

When compared to the PULSE results, the SoC arm of the model overestimated the number of CV-related deaths (and underestimated the number of non-CV deaths) when the >65-year-old subgroup

from EMPEROR-R was modelled. In PULSE, there were 7,905 CV deaths and 9,599 non-CV deaths over a mean follow-up of 3 years. In the model, there were [redacted] CV deaths and [redacted] non-CV deaths when the subgroup data from EMPEROR-R is used (for the first 3 years in the model).

The ERG requested that the company provided the KM data for all-cause and CV mortality in the above 65 years group and that, given the availability of KM CV and non-CV mortality data from PULSE, the KM curves from EMPEROR-R were adjusted to reflect a lower rate of death for the UK population subgroup. The ERG notes that this request was based on the company's model structure and the lack of KCCQ-CSS data from PULSE, which meant that the company could not use mortality data from PULSE directly in the SoC arm of the model. However, with the change in the company's approach to modelling survival in the model (see issue 7), the use of PULSE OS data directly in the model would have been possible, as the company included an option to estimate survival independently from KCCQ-CSS states in the updated model.

As a response to TE, the company reiterated its view that the difference in outcomes (i.e., HHF and CV-mortality rates) observed in EMPEROR-R and PULSE is likely due to inaccurate recording of events in PULSE rather than a clinically meaningful difference in the patients' characteristics. As discussed in Section 3.1, the ERG reiterates its view that PULSE remains the best available source of evidence to validate HF outcomes in the UK population.

The company ran a joint regression model of the individual patient-level data from EMPEROR-R and PULSE, including a term in the model for the population source (PULSE vs EMPEROR-Placebo) for all-cause and CV mortality with a Cox model. The estimated HRs represent the relative differences in mortality for HFrEF patients in the two sources and resulted in estimates of HR of [redacted] for all-cause mortality and of [redacted] for CV mortality. The company applied the HRs in the model to calibrate the survival results to the PULSE survival.

When compared to PULSE (7,905 CV deaths and 9,599 non-CV deaths over a mean follow-up of 3 years), the adjusted model predicted [redacted] CV deaths and [redacted] non-CV deaths when the subgroup data from EMPEROR-R is used (for the first 3 years in the model).

Given that the main source of benefit in the company's original model was the reduction in HHF events (and never the reduction in CV-related deaths), this scenario extended patients' life-years and resulted in a higher benefit associated with the number of avoided HHFs with empagliflozin which resulted in a small reduction in the final ICER.

3.9 Issue 9: Impact of hospitalisation for heart failure in patients' quality of life

The company clarified that the utility equation included indicators for time since the hospitalisation rather than duration. The ERG agreed with the company during the TE discussions, however noted that the company's approach was still likely to overestimate the impact of HHF on patients' QoL in the model given that it assumed that every HHF event would impact patients' quality of life for 1 year after the event.

The ERG notes that the [redacted] disutility per HHF event applied as a one-off in every month of the model was obtained by multiplying the coefficients for time since event (<1 month; 1 to <2 months; 2 to <4 months; 4 to <12 months) and adding these together to estimate an annual disutility [redacted]). The ERG discounted the disutility values for month 2;3;4;5;6;7;8;9;10;11 and 12 as these were applied as future events in every cycle of the model. The impact of adjusting the [redacted] disutility was negligible.

The ERG reinforces the fact that the second biggest driver of the QALY gains in the model comes from the reduction in HHF events associated with empagliflozin.

In the limited time available, the ERG searched the available literature to obtain more information on the impact of HHF on patients' quality of life. The dapagliflozin STA also assumed that HHFs impact patients' QoL over 12 months, with a [redacted] disutility of 0.32 per event. The sacubitril STA assumed that HHFs impact patients' QoL over 3 months, with a disutility of 0.213. For inclusiveness, the ERG has conducted a scenario analysis where the disutility assumed in the model was 0.213, however, the ERG recommends that the committee validates the company's assumption that 100% of HHFs impact patients' QoL for 12 months after the event.

3.10 Issue 10: Quality of life regressions for the UK population

During TE, the ERG noted that when the model was run for the >65 years old population, the baseline characteristics from the older EMPEROR-R subgroup were used in the QoL regression analysis, however the QoL regression was not re-estimated in this subgroup and thus the coefficients for the QoL predictors remained the same as those for the ITT population.

The company re-ran the regression analyses in the subgroup of patients aged 65 years or older. The coefficients from equations from the ITT population and 65+ subgroup were described in Table 17 of

the company's response to TE. Using effects derived from the subgroup equation in the economic model had a negligible impact on the final ICER.

3.11 Issue 11: Sex distribution underlying utility estimates

The ERG was concerned that the utility value associated with the KCCQ-CSS quartile 4 health state did not reflect the sex distribution in EMPEROR-R or PULSE. Therefore, the ERG suggested that the company adjusted the utility values in the model to reflect the sex distribution in the respective studies. The company carried the analysis and concluded that the change had minimal impact on utility. The ERG agrees with the company.

3.12 Issue 12: Quality of life gains in EMPEROR-Reduced

The ERG was concerned that the EQ-5D data from EMPEROR-R did not suggest a significant improvement in empagliflozin patients' QoL when compared to placebo patients and that the economic model generated a QALY gain of 0.14 (which remained unchanged in the company's model after TE) for the trial population.

The two main drivers of QALY gain in the model are related to: 1) how much longer empagliflozin patients stay in the better KCCQ-CSS states; and 2) the reduction in HHF experienced by empagliflozin patients. Results from the ERG's combined analysis in Section 4 show a QALY gain between 0.08 and 0.12 for empagliflozin in the trial population, depending on the assumptions made.

4 Results from ERG's exploratory analysis

In this section the ERG provides the results of the new exploratory analysis conducted after TE. Results of the exploratory analyses conducted using the trial population are reported in Table 9, while Table 11 reports the results in the UK population analysis. The following analyses were condcuted in both populations (as per the ERG's original analysis):

1. Assuming that 84% of patients receive lifelong treatment with empagliflozin;
2. Using the relative utility adjustment and the age-related decrements from Ara[1];
3. Replacing the proportion of UK patients who receive ACEi and ARNi in the model to reflect the ERG's clinical experts' opinion;
4. Using a unit cost for CV death in the model of £1,582;
5. Applying the ERG-calculated annual cost of dialysis (assuming 3 weekly sessions) of £23,088;

6. Assuming that HHFs impact patients' QoL over 3 months, with a disutility of 0.213.

In addition to the analyses above, the ERG also conducted the analyses listed below, which are specific to each population:

Trial population

a. Using the company's updated survival curves using a Weibull model and assuming that empagliflozin has no direct or indirect survival benefit over SoC. The ERG caveats this analysis by its' clinical experts' opinion that the long-term predictions of the Gompertz curves would be a better representation of the higher mortality seen in the trial population. Nonetheless, the company has only provided Weibull curves for the ITT population, therefore, the ERG could not use a Gompertz curve to estimate the impact of removing all the survival benefits associated with empagliflozin for the trial population.
b. Assuming that the proportion of patients in the KCCQ-CSS quartiles under the treatment arm is equal to those proportions in the placebo arm at 1 year + assuming no survival benefit for empagliflozin + assuming the TPS between KCCQ-CSS quartiles for patients on treatment are same as those for patients off treatment (after year 1).

UK population

c. Using the company's updated Weibull survival model and assuming that empagliflozin has no direct or indirect survival benefit over SoC.
d. Using the company's HR to adjust the survival to reflect PULSE survival.
e. Using the company's adjusted Poisson model to the PULSE HHF data.
f. Using the 0.723 utility value from Sullivan for the KCCQ-CSS quartile state (and adjusting other KCCQ-CSS state values accordingly)[2].
g. Assuming that the proportion of patients in the KCCQ-CSS quartiles under the treatment arm is equal to those proportions in the placebo arm at 1 year + assuming no survival benefit for empagliflozin (using the HR to adjust the survival to reflect PULSE) + assuming the TPS between KCCQ-CSS quartiles for patients on treatment are same as those for patients off treatment (after year 1).

Results in Table 9 show that the key driver of the economic results for the trial population is the assumption made for the effect of empagliflozin on patients' movements across KCCQ quartiles after year 1 in the model. The second driver of the economic results is the annual cost of dialysis.

As discussed in Issue 7, the decrease in the ICER when the survival benefit for empagliflozin is removed from the model is due to the increase in costs in the SoC arm being higher than the increase in costs in the empagliflozin arm, leading to a decrease of the incremental costs associated with empagliflozin when no survival benefit was assumed in the model. Overall, there was a decrease in the QALY gain associated with empagliflozin when the survival benefit was removed from the model which is likely to be due to the change in survival curves for the ITT population.

Table 9. Results of ERG's exploratory analysis in the trial population

Scenario		Incremental costs	Incremental QALYs	ICER
0	Company's base case post TE	£722	0.14	£4,999
1	Assuming that 84% of patients receive lifelong treatment with empagliflozin	£1,157	0.21	£5,465
2	Using the relative utility adjustment and the age-related decrements from Ara	£722	0.14	£5,208
3	Replacing the proportion of UK patients who receive ACEi and ARNi in the model to reflect the ERG's clinical experts' opinion	£710	0.14	£4,912
4	Using a unit cost for CV death in the model of £1,582	£733	0.14	£5,072
5	Applying the ERG-calculated annual cost of dialysis (assuming 3 weekly sessions) of £23,088	£815	0.14	£5,640
6	Assuming that HHFs impact patients' QoL over 3 months, with a disutility of 0.213.	£722	0.14	£5,211
a	Using the company's updated Weibull survival model and assuming that empagliflozin has no direct or indirect survival benefit over SoC.	£393	0.08	£4,777
b	Assuming that the proportion of patients in the KCCQ-CSS quartiles under the treatment arm is equal to those proportions in the placebo arm at 1 year + assuming no survival benefit for empagliflozin + assuming the TPS between KCCQ-CSS quartiles for patients on treatment	£446	0.05	£8,224

	are same as those for patients off treatment (after year 1).			

Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

The results of the combined exploratory analysis undertaken by the ERG are presented in Table 10. All the scenarios produced ICERs under the lower threshold of £20,000, even when it is assumed that empagliflozin had no effect on patients KCCQ-CSS scores after year 1 in the model.

Table 10. ERG's combined analysis in the trial population

Scenario		Incremental costs	Incremental QALYs	ICER
0	Company's base case post TE.	£722	0.14	£4,999
1	Assuming that 84% of patients receive lifelong treatment with empagliflozin.	£1,157	0.21	£5,465
1+2	Using the relative utility adjustment and the age-related decrements from Ara.	£1,157	0.20	£5,721
1+2+3	Replacing the proportion of UK patients who receive ACEi and ARNi in the model to reflect the ERG's clinical experts' opinion.	£1,139	0.20	£5,633
1+2+3+4	Using a unit cost for CV death in the model of £1,582.	£1,155	0.20	£5,711
1+2+3+4+5	Applying the ERG-calculated annual cost of dialysis (assuming 3 weekly sessions) of £23,088.	£1,295	0.20	£6,407
1+2+3+4+5+a	Using the company's updated Weibull survival model and assuming that empagliflozin has no direct or indirect survival benefit over SoC.	£839	0.12	£6,849
1+2+3+4+5+a+b	Assuming that the proportion of patients in the KCCQ-CSS quartiles under the treatment arm is equal to those proportions in the placebo arm at 1 year + assuming no survival benefit for empagliflozin + assuming the TPS between KCCQ-CSS quartiles for patients on treatment are same as those for patients off treatment (after year 1).	£926	0.09	£10,834

1+2+3+4+5+a+b+6	Assuming that HHFs impact patients' QoL over 3 months, with a disutility of 0.213.	£926	0.08	£12,234

Abbreviations: ICER, incremental cost-effectiveness ratio; HR, hazard ratio; PFS, progression free survival; PPS, post-progression survival; QALY, quality adjusted life year

Results in Table 11 show that the key driver of the economic results for the trial population is the impact of empagliflozin on patients' movements across KCCQ quartiles after year 1 in the model (once no survival benefit is assumed and the CV survival curves have been adjusted to reflect the PULSE population), followed by reducing the overall rate of HHF events in the model.

Table 11. Results of ERG's exploratory analysis in the trial population

Scenario		Incremental costs	Incremental QALYs	ICER
0	Company's base case post TE.	£971	0.13	£7,270
1	Assuming that 84% of patients receive lifelong treatment with empagliflozin.	£1,490	0.18	£8,089
2	Using the relative utility adjustmen.t and the age-related decrements from Ara	£971	0.13	£7,620
3	Replacing the proportion of UK patients who receive ACEi and ARNi in the model to reflect the ERG's clinical experts' opinion.	£958	0.13	£7,173
4	Using a unit cost for CV death in the model of £1,582.	£984	0.13	£7,368
5	Applying the ERG-calculated annual cost of dialysis (assuming 3 weekly sessions) of £23,088.	£1,051	0.13	£7,872
6	Assuming that HHFs impact patients' QoL over 3 months, with a disutility of 0.213.	£971	0.13	£7,460
c	Using the company's updated Weibull survival model and assuming that empagliflozin has no direct or indirect survival benefit over SoC.	£607	0.06	£9,780
d	Using the company's HR to adjust the survival to reflect PULSE survival.	£974	0.15	£6,708
e	Using the company's adjusted Poisson model to the PULSE HHF data.	£1,152	0.12	£9,678

f	Using the 0.723 utility value from Sullivan for the KCCQ-CSS quartile state (and adjusting other KCCQ-CSS state values accordingly).	£971	0.13	£7,601
g	Assuming that the proportion of patients in the KCCQ-CSS quartiles under the treatment arm is equal to those proportions in the placebo arm at 1 year + assuming no survival benefit for empagliflozin (using the HR to adjust the survival to reflect PULSE) + assuming the TPS between KCCQ-CSS quartiles for patients on treatment are same as those for patients off treatment (after year 1).	£636	0.04	£15,647

Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

The results of the combined exploratory analysis undertaken by the ERG are presented in Table 12. Most of the combined scenarios produced ICERs below the £30,000 threshold, with the exception of when it is assumed that empagliflozin has no effect on patients KCCQ-CSS scores after year 1 in the model, where the ICER increased to £48,767 per QALY gained. The ERG acknowledges this is an extreme scenario, where the benefit associated with empagliflozin on KCCQ-CSS outcomes is capped after year 1 in the model and is not likely to reflect clinical practice. This scenario served the purpose of exploring the uncertainty around the company's assumption that the changes seen in patients' KCCQ-CSS in EMPEROR-R from month 8 - 13 are sustained for approximately 30 years in the model.

The ERG conducted an additional scenario where the benefit associated with empagliflozin on KCCQ-CSS outcomes was capped after year 2 in the model and the ICER decreased from £48,767 to £26,000 per QALY gained. Therefore, the ERG concludes that the uncertainty around this parameter is unlikely to result in the "true" ICER being over £30,000.

When the ERG varied the disutility associated with HHF events in the model the ICER increased to £24,663 per QALY gained.

Table 12. ERG's combined analysis in the trial population

Scenario		Incremental costs	Incremental QALYs	ICER
0	Company's base case post TE.	£971	0.13	£7,270
1	Assuming that 84% of patients receive lifelong treatment with empagliflozin.	£1,490	0.18	£8,089

1+2	Using the relative utility adjustment and the age-related decrements from Ara.	£1,490	0.17	£8,525
1+2+3	Replacing the proportion of UK patients who receive ACEi and ARNi in the model to reflect the ERG's clinical experts' opinion.	£1,472	0.17	£8,425
1+2+3+4	Using a unit cost for CV death in the model of £1,582.	£1,491	0.17	£8,534
1+2+3+4+5	Applying the ERG-calculated annual cost of dialysis (assuming 3 weekly sessions) of £23,088.	£1,609	0.17	£9,211
1+2+3+4+5+c+d	Using the company's updated Weibull survival model and assuming that empagliflozin has no direct or indirect survival benefit over SoC. + using the company's HR to adjust the survival to reflect PULSE survival.	£1,317	0.11	£11,799
1+2+3+4+5+c+d+e	Using the company's adjusted Poisson model to the PULSE HHF data.	£1,748	0.08	£22,659
1+2+3+4+5+c+d+e +f	Using the 0.723 utility value from Sullivan for the KCCQ-CSS quartile state (and adjusting other KCCQ-CSS state values accordingly).	£1,748	0.07	£23,508
1+2+3+4+5+c+d+e +g	Assuming that the proportion of patients in the KCCQ-CSS quartiles under the treatment arm is equal to those proportions in the placebo arm at 1 year + assuming no survival benefit for empagliflozin (using the HR to adjust the survival to reflect PULSE) + assuming the TPS between KCCQ-CSS quartiles for patients on treatment are same as those for patients off treatment (after year 1).	£1,796	0.04	£48,767
1+2+3+4+5+c+d+e+6	Assuming that HHFs impact patients' QoL over 3 months, with a disutility of 0.213.	£1,748	0.07	£24,663

Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

5 Conclusion

The two key areas of uncertainty raised by the ERG before TE in the economic analysis where the long-term effect of empagliflozin on patients' change in KCCQ-CSS (in both the trial and in the UK population analyses sets) and the lack of representativeness of the subgroup data from EMPEROR-R when trying to replicate the UK population.

The ERG considers that the first area of uncertainty was resolved with the company's additional scenario analyses provided at TE. The ERG also notes that the ICER for empagliflozin compared to SoC is likely to remain under the £30,000 threshold in the trial population.

Nonetheless, the ERG remains concerned that the cost-effectiveness of empagliflozin compared to SoC in the UK population remains somewhat uncertain. Even though the cumulative scenarios presented by the ERG in this population demonstrate that the ICER for this population is likely to remain under the £30,000 threshold, the ERG remains uncertain if HHFs are accurately estimated in the long-term model.

The ERG maintains its view that using the KM HHF data from EMPEROR-R (independent from KCCQ-CSS states) would have allowed the company to estimate long-term HHF by relying on observed data and not assuming a constant rate of HHF. Importantly, the use of KM data would have allowed the company to model time to first and subsequent HHF separately and that this could be of importance given the results seen in EMPEROR-R for the ITT population.

Finally, the ERG notes that the clinical effectiveness of empagliflozin vs dapagliflozin is based on the results of the Bucher ITC, which are highly uncertain. Therefore, the ERG has provided a range of analyses (from assuming equivalence to assuming the mean estimates of the ITC accurately reflect the difference in treatments).

6 References

1. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value Health* 2010; **13**: 509-18.
2. Sullivan PW, Slejko JF, Sculpher MJ, Ghushchyan V. Catalogue of EQ-5D scores for the United Kingdom. *Med Decis Making* 2011; **31**: 800-4.