Everolimus (Certican®) for preventing organ rejection in liver transplantation

A critique of the submission from Novartis

Matrix and Matrix Knowledge are trading names of TMKG Limited (registered in England and Wales under registration number 07722300) and its subsidiaries: Matrix Decisions Limited (registered in England and Wales under registration number 07610972); Matrix Insight Limited (registered in England and Wales under registration number 06000446); Matrix Evidence Limited (registered in England and Wales under registration number 07538753); Matrix Observations Limited (registered in England and Wales under registration number 05710927); and Matrix Knowledge Group International Inc. (registered in Maryland, USA under registration number D12395794).

Disclaimer In keeping with our values of integrity and excellence, Matrix has taken reasonable professional care in the preparation of this document. Although Matrix has made reasonable efforts, we cannot guarantee absolute accuracy or completeness of information/data submitted, nor do we accept responsibility for recommendations that may have been omitted due to particular or exceptional conditions and circumstances.

Confidentiality This document contains information which is proprietary to Matrix and may not be disclosed to third parties without prior agreement.
Except where permitted under the provisions of confidentiality above, this document may not be reproduced, retained or stored beyond the period of validity, or transmitted in whole, or in part, without Matrix's prior, written permission.

© TMKG Ltd, 2012

Any enquiries about this report should be directed to enquiries@matrixknowledge.com

Report commissioned by:

NHS R&D HTA Programme

On behalf of:

NICE

Produced by:

Matrix Knowledge
Peninsula Technology Assessment Group (PenTAG)

Although Matrix are primarily responsible for the work in this report, PenTAG retains responsibility for the standard of the report and the quality of the advice that it contains.

Authors:

Mariana Bacelar, Economist, Optimity Matrix
Mitesh Nakum, Assistant Economist, Optimity Matrix
Adeline Durand, Senior Economist, Optimity Matrix
Chris Cooper, Senior Information Specialist, PenTAG
Chris Hyde, Professor of Public Health and Clinical Epidemiology, PenTAG
Nick Latimer, Research Fellow in Health Economics, University of Sheffield
Jaime Peters, Research Fellow, PenTAG
Ruben Mujica-Mota, Senior Lecturer, PenTAG

Acknowledgements:

We particularly acknowledge the help of James O'Beirne, Royal Free Hospital who advised us.

Declaration of competing interest of the authors

None

Rider of responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows

Bacelar M, Nakum M, Durand A, Cooper C, Mujica-Mota R, Hyde C, Latimer N, Peters J. Everolimus (Certican®) for preventing organ rejection in liver transplantation. Single Technology Appraisal NIHR HTA Programme, Matrix and Peninsula Technology Assessment Group, 2014.

Contribution of authors

Mariana Bacelar: Contributed to project management, the critique of the company's submission, report writing and editing.

Mitesh Nakum: Contributed to the critique of the company's submission, report writing and editing.

Adeline Durand: Commented on drafts of the report.

Chris Cooper: Commented on the searches provided by the company and contributed to report writing.

Chris Hyde: Contributed to the critique of the company's submission and commented on drafts of the report.

Nick Latimer: Contributed to the critique of the company's submission.

Jaime Peters: Contributed to the critique of the company's NMA.

Ruben Mujica-Mota: Contributed to the critique of the company's submission and commented on drafts of the report and is the guarantor of the report from PenTAG.

Correspondence to:

Optimity Matrix
1st Floor | Kemp House | 152-160 City Road | London | EC1V 2NP

Date completed:

November 2014

Contents

1.0 Summary 11
1.1 Scope of the submission 11
1.2 Summary of submitted clinical effectiveness evidence 11
1.3 Summary of submitted cost effectiveness evidence 11
1.4 Commentary on the robustness of submitted evidence 12
1.5 Key issues 12
2.0 Background 14
2.1 Critique of company's description of underlying health problem 14
2.2 Critique of company's overview of current pathway of care and service provision 14
3.0 Critique of company's definition of decision problem 16
3.1 Population 16
3.2 Intervention 16
3.3 Comparators 17
3.4 Outcomes 20
3.5 Time frame 21
4.0 Clinical effectiveness 21
4.1 Critique of company's approach 21
4.2 Summary of submitted evidence 41
5.0 Economic evaluation 60
5.1 Overview of company's economic evaluation 60
5.2 Critique of approach used 86
5.3 Results included in company's submission 128
6.0 Additional work undertaken by the ERG 135
6.1 Correction for errors in Novartis' model 135
7.0 Summary of clinical and cost-effectiveness issues 137
8.0 References 140

List of tables

Table 1. Updated incidence of liver transplants in the UK 14
Table 2. Drug doses in comparator studies 18
Table 3. Eligibility criteria used in the search for RCTs..... 23
Table 4. Details of the H2304 Study 25
Table 5. Excluded studies containing a regimen of everolimus in combination with tacrolimus. 26
Table 6. Clinical appraisal of relevant RCTs 29
Table 7.Description of the 16 endpoints considered for the NMA..... 33
Table 8: Model fit statistics 39
Table 9. KM incidence rates of primary efficacy endpoint (ITT population at 12 months) 41
Table 10. KM incidence rate of graft loss or death (ITT population at 12 months) 42
Table 11. Incidence rates of secondary efficacy endpoints (ITT population at 12 months)..... 42
Table 12. KM incidence rates of primary efficacy endpoint (ITT population at 24 months) 43
Table 13. KM incidence rates of graft loss or death (ITT population at 24 months) 44
Table 14. Incidence rates of secondary efficacy endpoints (ITT population at 24 months)..... 44
Table 15. KM incidence rates of efficacy endpoints at Month 36 (ITT population) 45
Table 16. eGFR (ITT population at 12 months) 49
Table 17. eGFR (ITT population at 24 months) 49
Table 18. Expected absolute results for renal function at 12 months (fixed effects model) 56
Table 19. Key adverse events across randomised groups from H2304 58
Table 20. Health states considered in the hepatic core model 62
Table 21. Health states in renal sub-model..... 63
Table 22. Three-month transition probabilities used in the base case hepatic model across treatment arms 67
Table 23. Absolute eGFR decrease at 12 months as used in the renal sub-model 71
Table 24. Expected absolute results for renal function at 12 months (NMA fixed effects model) 71
Table 25. Transition probabilities in the renal sub-model independent of treatment effectiveness 72
Table 26. Hepatic utilities used in the excel model 74
Table 27. Weighted average of QoL data – Novartis analysis to estimate utility in the SPT state 75
Table 28. Renal utilities used in the excel renal sub-model..... 76
Table 29. Incidence of AEs in the economic model 77
Table 30. AEs disutility data..... 78
Table 31. AEs cost data 78
Table 32. Health state costs used in the hepatic model 79
Table 33. Health state costs used in the renal sub-model 82
Table 34. Intervention and comparators' costs 84
Table 35. Critical appraisal checklist based on NICE Reference Case (NICE, 2008) 87
Table 36. Critical appraisal checklist from Drummond and colleagues (Drummond et al. 1997)88
Table 37. Critical appraisal checklist of Philips et al (2004) for model-based analysis..... 89
Table 38. Baseline caracteristics varied in the patient simulation model..... 93
Table 39. Example of allocation of patients to health states in EVR+rTAC arm of Novartis model 99
Table 40. Probability of AR used in the economic model 99
Table 41. Simulations ran by the ERG – no changes to the company model 102
Table 42. Baseline population in the economic model – all treatment arms..... 105
Table 43. TAC trough levels in the NMA studies. 114
Table 44. CKD progression rates used by Novartis 116
Table 45. CKD baseline values and progression probabilities used in La Mattina. 116
Table 46. CKD increased risk of mortality (over 3 months) used by Novartis 119

Table 47. CKD mortality over 5 years reported in Keith et al, 2014. 119
Table 48. Post-transplantation ongoing care costs in the renal model 126
Table 49. Mean daily dose of immunosuppression assumed in the model 127
Table 50. Base case ICERs 129
Table 51. Average time spent in different health states of the economic model 130
Table 52. Scenario 1 run by Novartis 132
Table 53. Scenario 2 run by Novartis 133
Table 54. Scenario 4 run by Novartis 133

List of figures

Figure 1. Kaplan-Meier plot of composite efficacy failure at Month 36 since randomization 46
Figure 2. KM for the proportion of patients free from graft loss or death (ITT population -36 months) 47
Figure 3. Subgroup analysis at 12 months 48
Figure 4. Expected absolute results for overall survival at 12 and 24 months (fixed effects model) 52
Figure 5. Expected absolute results for graft survival at 12 and 24 months (fixed effects models) 53
Figure 6. BPAR free at 3, 6 and 12 months (fixed effect models) 55
Figure 7. Network for renal function at 12 months 57
Figure 8. Core hepatic cost-effectiveness model and renal sub-model 62
Figure 9. NMA tBPAR estimates used for transition probabilities in the hepatic model 66
Figure 10. Baseline population in H2304 105
Figure 11. TAC trough levels by visit – local lab data, safety population – 12 months 107
Figure 12. Proportion of patients alive in the economic model 121

LIST OF ABBREVIATIONS

Abbreviation	Definition
AE	Adverse Events
ANCOVA	Analysis of Covariance
ATG	Anti-lymphocyte globulin
AZA	Azathioprine
BPAR	Biopsy proven acute rejection
BTS	British Thoracic Society
CCR	Creatinine Clearance Rate
CKD	Chronic Kidney Disease
CIC	Ciclosporin
CNI	Calcineurin inhibitor
CRD	Centre for Reviews and Dissemination
DIC	Deviation Information Criterion
DoH	Department of Health
DSU	Decision Support Unit
eGFR	Estimated glomerular filtration rate
EMA	European Medicines Agency
EQ-5D	Euroqol 5-dimensions
ESLD	End Stage Liver Disease
ETOH	Alcoholic liver disease
EVR	Everolimus
HCC	Hepatocellular Carcinoma
HCV	Hepatitis C Virus
HIV	Human Immunodeficiency Virus
HR	Hazard Rate
HRG	Health Resource Group
HRQL	Health Related Quality of Life
HSUV	Health State Utility Values
HTA	Health Technology Assessment
ICER	Incremental Cost Effectiveness Ratio

ILTS	International Liver Transplantation Society
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
ITC	Indirect Treatment Comparison
ITT	Intention to Treat
IVRS	Interactive Voice Response System
KM	Kaplan Meier
LY	Life year
MCMC	Markov Chain Monte Carlo
MDRD4	Four-variable Modification of diet in renal disease
MELD	Model for End-Stage Liver Disease
MMF	Mycophenolate Mofetil
MTC	Mixed Treatment Comparison
mTOR	Mammalian target of rapamycin
NASH	Non-Alcoholic Steato-hepatitis
NCGC	NICE Clinical Guideline Centre
ng/ml	Nanogram per millilitre
NHS EED	NHS Economic Evaluation Database
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMA	Network Meta-Analysis
NR	Not Reported
PICOS	Population Intervention(s)/Comparators(s), Outcome(s) and Study Design
PSSRU	Personal and Social Services Research Unit
QALY	Quality Adjusted Life Year
RCT	Randomised Controlled Trial
rTAC	Reduced dose Tacrolimus
SD	Standard Deviation
SE	Standard Error
SIGN	Scottish Intercollegiate Guidelines Network
SLR	Systematic Literature Review
SMC	Scottish Medicines Consortium
SPC	Summary of product characteristics

tBPAR	Treated biopsy proven acute rejection
TAC	Tacrolimus

1.0 Summary

Text cited directly from the submission by Novartis (hereafter referred to as 'the submission') is presented in italic and cross referenced.

1.1 Scope of the submission

The submission from Novartis considered the use of everolimus (Certican®) in combination with reduced dose tacrolimus with corticosteroids for the initial 6 months of therapy, for adults on maintenance immunosuppressive therapy following allogeneic liver transplantation. The comparators considered are tacrolimus in combination with mycophenolate mofetil and azathioprine in combination with tacrolimus, both with corticosteroids for the initial 6 months of therapy.

1.2 Summary of submitted clinical effectiveness evidence

EVR+rTAC effectiveness relied on evidence drawn from the H2304 trial, which is of good quality. The trial included three treatment arms, an EVR+rTAC arm, a standard TAC arm and a TAC discontinuation arm which was discontinued. It seems that the composite primary endpoint of the trial (treated biopsy-proven acute rejection (tBPAR), graft loss and death) combines two outcomes on which EVR+rTAC has effects in different directions relative to the comparator. Whilst graft loss or death seem to favour the TAC arm of the trial, the addition of tBPAR to the composite endpoint favours the EVR+rTAC arm. This casts doubt on the appropriateness of the clinical endpoints as previously discussed.

The trough levels of reduced TAC in the H2304 trial are above of what would be considered as reduced levels in the UK clinical practice.

The ERG found a significant limitation in the Network Meta-Analysis (NMA) presented by Novartis as the data included in the WinBUGS codes do not relate to the submission data in any way. This poses a major limitation as the ERG could not verify which data were used for the analysis of specific outcomes. This represents a concern as the ERG are not clear which studies have been included in the NMA analysis for the tBPAR outcome, due to lack of clarity and transparency in the submission.

The scenario analysis undertaken for the NMA consisted in removing specific trials from the network to assess the impact on the NMA results. Nonetheless this analysis lacks transparency and is not informative.

1.3 Summary of submitted cost effectiveness evidence

Novartis developed a cost-utility model as patient simulation model. The model includes a hepatic model and a renal sub-model.

Novartis' model produced ICERs for EVR+rTAC compared to AZA+TAC of £187,842 and an ICER of £110,797 for EVR+rTAC compared with MMF+TAC.

1.4 Commentary on the robustness of submitted evidence

1.4.1 Strengths

It is likely that the systematic review on the literature for clinical effectiveness, QoL, AE and cost data undertaken by Novartis contains all the relevant studies. The evidence submitted generally reflect the decision problem outlined in the final scope of the submission.
The H2304 RCT is of good quality.

1.4.2 Weaknesses

The ERG are extremely concerned with the **lack of stability in the base case ICERs.** In particular the ERG could not replicate the base care results reported in the submission based on the information contained therein.

The ERG also have concerns regarding the **validity of the NMA**, the **model structure** and the **clinical inputs** used by Novartis.

The ERG lack confidence in the final ICER presented. Novartis' model produced ICERs for EVR+rTAC compared to AZA+TAC of £187,842 and an ICER of £110,797 for EVR+rTAC compared with MMF+TAC. Furthermore the undertaken probabilistic sensitivity analysis lacks robustness.

Finally, the fact that the EVR+rTAC, the MM+TAC and AZA+TAC arms are not reflective of clinical reality in the UK in terms of TAC levels is worrying as it questions the validity and the applicability of the economic analysis within the NHS context.

For all the reasons provided, the ERG lack confidence in the final ICER presented and in the company analyses and therefore cannot make any predictions regarding the correct final ICERs.

1.5 Key issues

The ERG are extremely concerned with the **lack of stability in the base case ICERs**. The ERG ran two iterations of the base case model as submitted by the company, using the number of recommended simulations and **without any changes,** to test the model stability with regards to ICER results. The ERG found a considerable variation in the ICERs reported, especially in the case of the ICERs for EVR+rTAC versus MMF+TAC, which varied between £110,797 and £120,651 (nearly a 9% change).

The ERG are overall concerned with the **validity of the NMA** undertaken by Novartis. The approach undertaken raises the following concerns:

- The ERG are not clear what were the final studies included in the tBPAR analysis.
- Overall, the ERG question the validity of the NMA results for the renal outcomes. The allocation of the different studies' treatment arms to the reduced and standard TAC categories

is inconsistent and misleading. As the standard TAC connector across the NMA studies is so heterogeneous the NMA results are likely to lack robustness.

The ERG are overall concerned with the **model structure and assumptions** used by Novartis. The approach undertaken raises the following concerns:

- The ERG are not convinced that a patient simulation state-transition model is required.

- The ERG believe that more emphasis should have been placed on the renal component of the economic model. More specifically, interaction between the 2 models should have been considered as it is known that not only immunosuppressive therapy for liver post-transplantation has an impact on renal functioning but also, equally important, that renal functioning has an impact on graft survival.

- We question the clinical plausibility of the mild chronic rejection (MCR) state in the model and especially the high costs associated with this state. Removing this state from the economic analysis led to an increase in the ICER for EVR+rTAC compared with MMF+TAC from £110,797 to £227,528, whilst the ICER comparing EVR+rTAC with AZA+TAC decreased from £187,842 to £176,410.

- External validation of the model outputs showed that overall survival is underestimated in the economic model.

- The way the simulation was run in the model produces clinical implausible results in the renal model. Due to the random generation process used to estimate the baseline eGFR levels, some patients start the model with negative levels of eGFR, which is not clinically realistic.

- We have identified errors in the formulae allocating patients to different health states in the model which:

    - Are likely to have been avoided if cohort state-transition model had been used.
    - Could not be corrected by the ERG due to the computational burden of running the simulation model.
    - Are likely to bias the results in favour of EVR+rTAC in the hepatic model.

The ERG are worried with the **clinical inputs** used by Novartis. The ERG found that some of the parameters being used in the economic model differ from the ones reported in the submission and that some assumptions made with regards to resource use and drug effectiveness on renal outcomes should be further tested and validated.

2.0 Background

2.1 Critique of company's description of underlying health problem

In Sections 2.1 to 2.3 of their submission, Novartis describe the underlying health problem. They provide a summary of the characteristics and progression of end stage liver disease (ESLD). It is mentioned how the most common underlying causes of ESLD in Europe are hepatitis C virus infection (HCV), nutritional-toxic liver cirrhosis (NTLC) and hepatocellular carcinoma (HCC). It is also explained how patients with ESLD rely on liver transplantation to survive as there are is no suitable alternative treatment for these patients.

Superseded
See Erratum

The complications associated with liver transplantation are summarized. The company explains that in the short-term there is a risk of acute rejection of the graft which can progress to chronic rejection if the immunosuppressive therapy is not adequately managed. Long-term complications are typically associated with the recurrence of the underlying liver disease or the immunosuppressive regimen used (these typically include renal dysfunction, new onset of diabetes mellitus and cardio-vascular disease).

The company provide information on the median survival time of adult liver allograft recipients in the UK (based on Barber et al, 2007). The estimated median survival is 22.2 years (CI 19.3 - 25.6, p-value 0.05) and an estimated loss of 7 years compared with age and sex matched population. The median survival for specific age groups is also provided, with patients between 24 and 54 years reporting a median survival of 25.3 years (CI 20.5–31.2) and 55 to 64 year old patients, reporting a median survival of 19.5 years (CI 16.1–23.6).

To note is that Barber et al, 2007 also reports the median survival for transplant patients with specific underlying liver disease. Patients with alcoholic, HCV and cancer liver disease had a median survival of 15, 12 and 5.3 years, respectively.

Novartis present the incidence of liver transplants from April 2012 – March 2013 based on a reliable relevant source (NHS Blood and Transplant). However, the NHS Blood and Transplant present more recent figures in their website from 1 April 2013 to 31 March 2014, which the ERG reproduce here.

Table 1. Updated incidence of liver transplants in the UK

Description	UK estimate	Source
Liver transplants in the UK (including deceased and living transplants) April 2013 – March 2014	881	http://www.organdonation.nhs.uk

2.2 Critique of company's overview of current pathway of care and service provision

Novartis mention that existing NICE guidance and national protocols do not cover the management of immunosuppressive treatment for patients who undertook liver transplant.

However the ERG found the British Transplantation Society (BTS) Guidelines for Liver Transplantation for Patients with Non-Alcoholic Steato-Hepatitis (NASH). Novartis mention the document on page 138 of their submission, therefore it is not clear to the ERG why the company considered that *no single guideline was identified for the immunosuppression of adult patients with allogeneic liver transplantation.* Furthermore, on page 135 of their submission Novartis explain that the economic analysis considered patients transplanted for any kind of liver failure, such as NASH.

Clinical opinion sought by the ERG explained that there is huge variance in clinical practice with regards to post-transplantation immunosuppressive treatment regimens in the UK. It also agreed that the only available guidance on the management of immunosuppressive treatment for patients who undertook liver transplant is the BTS NASH guidance and deemed this extremely relevant for the context of the decision problem.

Superseded
See Erratum

The company explains that everolimus will fit in the existing post-transplantation clinical pathway as it offers an alternative to current immunosuppressive therapy. It is also mentioned that current practice shows variations with regards to choice of immunosuppressive therapy and long-term treatment to ensure graft survival whilst avoiding treatment-related complications.

On page 33, it is reported that calcinerium inhibitors (CNIs) remain the *backbone of immunosuppression* for post-transplanted liver patients regardless of the well-known long-term complications associated with these, such as renal toxicity. Therefore there seems to be a delicate balance between the use of these to reduce acute rejection and trying to reduce CNI-related complications in the long-term. Because everolimus acts synergistically with CNIs, there is an opportunity to minimise the CNI dose administered to patients after liver transplant, therefore sparing them from some of the complications associated with long-term use of CNIs.

The company point to the fact that in the UK clinical practice is based on tacrolimus-containing regimens for the prophylaxis of hepatic graft rejection.

There is not much consideration given to the main immunosuppressive therapies used in the UK apart from mentioning there is variance in practice and that most regimens contain tacrolimus in combination with other drugs. However, for example, the NASH guidance mention tacrolimus monotherapy and sirolimus as relevant therapies.

Clinical opinion sought by the ERG advised that about 10-15% of patients currently receive sirolimus, which like everolimus is an mTOR inhibitor, instead of everolimus during their first year of immunosuppressive therapy. Use of sirolimus is significantly lower than everolimus since the drug is poorly tolerated and has been known to cause hepatic artery thrombosis. Everolimus is much better tolerated, and can be introduced in the immunosuppressive therapy earlier than sirolimus.

Furthermore, the ERG's clinical advisor explained that around 30% of patients receive tacrolimus monotherapy, which will be adjusted to lower doses as the patient gets stable and that around 70% of patients receive therapies including 2 agents, like mycophenolate acid or azathioprine in combination with reduced doses of tacrolimus.

The company also mention other drugs relevant to the liver post-transplant clinical pathway. These are steroids, in case of acute rejection, anti-viral/antibiotics for patients at risk of developing cytomegalovirus infection (CMV) or any bacterial infection.

It is noted how post-transplanted patients should be monitored for occurrence of hyperlipidaemia, new-onset diabetes and renal dysfunction. Patients with renal dysfunction are managed with a combination of pharmaceutical intervention and dialysis or renal transplantation for those patients who progress to end-stage renal disease.

3.0 Critique of company's definition of decision problem

3.1 Population

The population defined in the NICE Scope[1] is adults undergoing allogeneic liver transplantation.

The population considered by the submission is adults on maintenance immunosuppressive therapy following allogeneic liver transplantation.

The company explains that the expected license for everolimus is for prophylaxis of graft rejection after liver transplantation, therefore the submission only addresses the maintenance therapy period (starting at 4 weeks after transplantation).

Focusing on the period after transplantation seems appropriate given everolimus therapeutic indication and the EGR's clinical adviser's opinion.

3.2 Intervention

The intervention defined in the scope is Everolimus (EVR) in combination with tacrolimus (TAC) and a corticosteroid.

The intervention under assessment is everolimus (Certican®) in combination with reduced dose of tacrolimus (hereafter referred as EVR+rTAC) *with or without* corticosteroids. Everolimus is a protein kinase inhibitor, more specifically it is an inhibitor of the mammalian target of rapamycin (mTOR).

The intervention under assessment in the submission departs from the scope as it specifies a <u>reduced dose</u> of TAC and by considering EVR+rTAC <u>both with and without corticosteroids</u>. A reduced dose of TAC is appropriate for the purpose of the submission.

With regards to the use of corticosteroids, the ERG find the sentence "with or without corticosteroids" slightly vague. The economic analysis considered the use of corticosteroids for the initial 6 months of therapy, after which these were removed from all treatment regimens. Clinical opinion sought by the ERG agreed this to be appropriate and added that after 6 months of immunosuppressive therapy, around 90% patients will be on corticosteroids-free therapies.

[1] Referred to as "the scope" in the remainder of this report.

EVR was submitted to the MHRA on the 27th May 2014 and UK approval is anticipated for early December 2014. The anticipated UK indication for liver transplantation is f*or the prophylaxis of organ rejection in patients receiving a hepatic transplant. In liver transplantation, Certican should be used in combination with tacrolimus and corticosteroids.* The company also expect Certican to be assessed by the SMC, with a submission estimated for [redacted] and with guidance expected to be published in [redacted].

The company reports that the EVR recommended starting dose is 1.0 mg twice daily, 4 weeks after transplantation. The EVR dose is targeted to remain at a trough level of 3 to 8 ng/mL in combination with reduced TAC at 3 to 5 ng/mL and corticosteroids, which can be adjusted if necessary.

Whilst for EVR the reported doses are in line with the treatment regimen followed in the main clinical trial (H2304), source of effectiveness evidence submitted by the company, this is not the case for reduced TAC, where trough levels in the trial were above 5. Of note is that in the UK, a reduced dose of TAC is defined as blood through levels of TAC <5 ng/mL. This will be further discussed throughout the ERG report.

Superseded
See Erratum

The company does not specify the corticosteroid recommended dose, however in H2304 this was given at a minimum dose of 5mg per day for 6 months.

The company reports that dose adjustments might be necessary. The standard EVR dose should be halved for patients with moderate or severe hepatic impairment (Child-Pugh B or C).

Treatment is expected to be continued throughout the remaining of the graft/patient life.

3.3 Comparators

The comparator set out in the scope was defined as a standard immunosuppressive therapy with a CNI (such as ciclosporin or tacrolimus) and a corticosteroid in combination with azathioprine or mycophenolate acid.

Novartis report that the choice of relevant comparators was based on historical discussions with NICE and a scoping workshop (held on the 24th July 2013) which resulted in the selection of tacrolimus in combination with mycophenolate mofetil with or without corticosteroids (hereafter TAC+MMF) and azathioprine in combination with tacrolimus, with or without corticosteroids (hereafter TAC+AZA) as the relevant comparators.

Thus the comparator used was TAC+AZA and TAC+MMF, both with corticosteroids for the initial 6 months of therapy.

Therefore, the comparator used in the submission departs from the one set out in the initial scope in three aspects:

1. It excludes ciclosporin from the economic analysis
2. It considers mycophenolate mofetil (MMF) instead of mycophenolate acid
3. It considers drug regimens both with and without concomitant corticosteroids

Clinical opinion sought by the ERG indicated that approximately 7% of patients have ciclosporin included in their treatment regimen but that this drug would not be considered as a first-line treatment option due to its high rate of adverse events, therefore the exclusion of ciclosporin seems to be appropriate.

The scope originally considered mycophenolate acid. However the company decided to define mycophenolate mofetil as the comparator drug. Clinical opinion sought by the ERG explained that mycophenolate acid includes both mycophenolate sodium and mycophenolate mofetil. Even though the active component of mycophenolate mofetil and mycophenolate sodium is the same, these drugs are not interchangeable as 500mg of mycophenolate mofetil is equivalent to 360mg of mycophenolate sodium.

Furthermore the justification provided by Novartis on page 36 of their submission regarding the choice of mycophenolate mofetil is not clear to the ERG. Therefore it seems that mycophenolate acid should have been used as the comparator drug as it is less restrictive than mycophenolate mofetil or mycophenolate sodium.

Superseded
See Erratum

As previously mentioned, the exclusion of corticosteroids after 6 months is appropriate as only patients with specific co-morbidities such as autoimmune hepatitis will be kept on corticosteroids after 6 months (as per clinical opinion sought by the ERG).

Due to a wide range of studies considered for the NMA and time constraints, the ERG approach was to focus on:

- The studies included in the final NMA (Novartis needed to refine the studies included in the analysis due to incompatibility reasons)
- The studies used to derive the parameters included in the economic model (tBPAR and renal functioning)

Nevertheless, the ERG are still not clear which studies have been included in the final NMA analysis for the biopsy proven acute rejection (tBPAR) outcome, due to lack of clarity and transparency in the submission.

The studies included in the NMA to obtain renal outcomes were the H2304 trial, Neuberger et al, 2009, Boudjema et al, 2011, Porayko et al., 1994 and McDiarmid et al., 1993. The doses for the comparator drugs are presented in Table 2.

Table 2. Drug doses in comparator studies

(Neuberger *et al.*, 2009[65])	Standard TAC	The standard TAC arm presents average trough levels of 11 ng/mL until month 3 and then decreased to 9 by the end of the 1st year.
	MMF + reduced TAC	The reduced TAC arm presents average trough levels of 9 ng/mL until month 3 and then decreased to 8 by the end of the 1st year.

		MMF, 2g/day
(Boudjema *et al.*, 2011[83])	Standard TAC	The standard TAC arm presents average trough levels above 10 ng/mL for the whole of the 1st year. The reduced TAC arm presents average trough levels between 10 and 9 ng/mL for the 1st year. MMF, 1.5 g twice a day for the first 6 weeks, 1.0 g twice a day until month 12.
	MMF + reduced TAC	
(Porayko *et al.*, 1994[88])	Standard TAC	TAC trough levels were maintained below 1.0 ng/mL for the entire year. AZA: 2mg/kg/day.
	AZA+ciclosporin	
(McDiarmid *et al.*, 1993[93])	Standard TAC	TAC trough levels were maintained below 2 ng/mL.
	AZA+ciclosporin	AZA, 1-2mg/kg/day.

Superseded
See Erratum

There are three additional points which arose from discussions with our clinical expert which the ERG believe to be worthy of consideration:

- When evaluating the benefit of EVR as a concomitant drug prescribed with a reduced dose of TAC and with regards to its renal sparing effect, EVR is likely to perform as good as MMF and AZA when given with reduced doses of TAC. This is because the renal sparing effect comes from allowing reduced doses of TAC (versus standard doses of TAC), therefore using EVR, MMF or AZA is likely to lead to the same renal outcomes. In this case, the cheaper drug is likely to be the most cost-effective therapy. That would be AZA.

- The true advantage of EVR and the uniqueness of the drug is likely to be the fact that it can be used as a monotherapy regimen. The ERG are aware that the drug indication is not monotherapy, nonetheless we consider this to be an extremely relevant issue. Whilst MMF and AZA cannot be given as monotherapy regimens (as it has been shown this leads to an increase in graft loss), EVR seems to perform well when taken alone. Disappointingly, it seems that not enough data exist to show the benefit of EVR monotherapy. Even though the H2304 trail initially designed a TAC discontinuation arm, where patients were kept on EVR monotherapy, this arm was discontinued before the end of the trial, as it reported higher rates of episodes of acute rejections. However our clinical advisor explained that this is likely to be related to the inappropriate choice of clinical endpoints for the trial which focused on episodes of acute rejection instead of long-term survival of the graft. Nowadays it seem to be broadly accepted that patients can experience (up to 2) episodes of acute rejection without these impacting on the long-term survival of the graft as these are easily treated and have around 90% of successful resolution (also demonstrated in Novartis' economic analysis). Therefore, whilst the number of acute rejections is a relevant endpoint, long-term survival would be a better one for clinical trials evaluating the effectiveness of immunosuppressive therapies.

- These two issues beg the discussion of the appropriateness of the comparators used in the context of this decision problem. It can be argued that to be able to truly isolate the effect of

EVR in terms of its renal sparing effect, the most suited comparator would be reduced TAC. Not only TAC monotherapy is widely used in clinical practice (with reduced doses of TAC), but also this is the only way to isolate the effectiveness of EVR. The comparators used in this submission are unlikely to allow the true benefit of EVR to be assessed for two main reasons:

- The TAC dose is not reduced. There seems to be no justification for not using reduced doses of TAC as a comparator, even if given concomitantly with other drugs (like AZA and MMF) since this is widely used in clinical practice;

- The comparator regimens include AZA and MMF, which in a way work as confounding factors. Including these drugs in the analysis adds "noise" to it as the final outcomes would ideally reflect the true effectiveness of EVR vs TAC. When adding different comparators, the EVR renal sparing effect cannot be truly "isolated".

In conclusion the comparator set used make the analysis somewhat irrelevant as it does not include reduced TAC monotherapy as a comparator but also considers comparators that from a clinical point of a view are not very relevant (AZA and MMF given with standard doses of TAC which does not reflect clinical practice at later stages of immunosuppressive therapy).

Finally, the fact that the effectiveness of EVR+rTAC is based on the H2304 study, where the "reduced" TAC doses are actually the equivalent to what is considered standard doses of TAC in the UK, adds to the difficulty of understanding the true benefit of EVR.

Even though the recruitment to the TAC elimination arm was discontinued, some patients were allowed to remain on this arm of the trial and were followed up to 36 months. The ERG review some of these data throughout this report.

3.4 Outcomes

The outcomes considered in the submission include:

- Patient survival
- Graft survival
- Time to acute rejection
- ***Time to recurrence of hepatocellular carcinoma***
- Renal function
- Time to end-stage renal disease
- Adverse effects of treatment
- Health-related quality of life

This departs from the outcome measures considered in the scope, which included graft function and graft fibrosis. Upon the ERG's request for clarification the company reported a typographical error and clarified that time to recurrence of hepatocellular carcinoma (HCC) was in fact included as an outcome. Novartis clarified that expert opinion advised that *there was little evidence to differentiate between interventions with regards to* graft function and graft fibrosis. HCC was included in the searches as a relevant outcome however, after a feasibility analysis the company decided that it was not possible to include this outcome in the network meta-analysis for comparators.

Clinical opinion sought by the ERG informed that graft function and fibrosis are difficult to measure as their definition is not standardized. Additionally, the impact of immunosuppressive therapy on graft function and fibrosis would only become apparent around year 3 of treatment. The same applies to the recurrence of HCC, which would only become apparent after 2 years of therapy. Therefore the ERG's clinical advisor explained that the main clinical trial H2304 (which lasted for 1 year with a follow-up of 36 months) would not necessarily capture the impact of EVR+rTAC on these outcomes. However, it was mentioned that if any impact was to be observed it was likely to benefit EVR+rTAC (against standard TAC).

Superseded
See Erratum

3.5 Time frame

The time horizon considered in the economic model was lifetime (80 years). Given that the average staring age of patients in the economic model was 54 years (with a standard deviation of 10 years), the time frame seems unnecessarily high, adding to the computational burden of the model. Around year 40 of the analysis 100% of patients are virtually dead.

4.0 Clinical effectiveness

4.1 Critique of company's approach

In this chapter we assess the clinical evidence provided by Novartis in their submission.

We start with a description and critique of Novartis's literature search strategy, followed by a description of the main studies selected for clinical effectiveness and their quality assessment. We then look at the company's selection of outcomes and the statistical approach they used to analyse them. This is followed by a summary of their submitted evidence for clinical effectiveness and our comment on their validity.

4.1.1 Description of company's search strategy and comment on whether the search strategy was appropriate

- Clinical Effectiveness Searches

The ERG are happy to accept the clinical effectiveness searches as presented by the company.

Novartis ran four systematic literature reviews in order to identify relevant published and unpublished clinical data. These targeted:

1. Clinical data for the intervention of interest, EVR+rTAC
2. Studies suitable for a network meta-analysis (NMA), both for the intervention and the two comparator regimens (AZA+TAC and MMF+TAC) with or without corticosteroids
3. Non-RCT data for the intervention of interest, EVR+rTAC
4. Adverse event (AE) data

The search strategy was last updated in August 2014, so the results are considered current for this submission. The effectiveness search syntax took the following form:

(Terms for liver transplant or hepatic transplantation or graft) AND (terms for Everolimus (including brand names Certican or Zortress)) OR (terms for Azathioprine) OR (terms for mycophenolic acid) AND (terms for cyclosporine) OR (terms for tacrolimus).

Novartis searched all of the required bibliographic databases, in addition to clinical trial registries and conference proceedings. The ERG are content with the range of resources used in this submission and, therefore, the company's attempts to locate published and unpublished RCT evidence.

Superseded
See Erratum

The ERG point to the following limitations of the searches undertaken:

The search returns are limited to studies that use cyclosporine or TAC in combination with EVR or AZA or MMF. Any studies that evaluate the efficacy of EVR as a stand-alone intervention, would be missed by this search.

The bibliographic searches were date limited 1990-Current and the conference proceedings were date limited 2012-current.

The ERG noticed a small typographical error in the clinical effectiveness searches. The Boolean connecter OR had been inadvertently omitted between mycophenolic acid / morpholinoethyl ester. The line was presented in the company's submission as:

(cellcept or mycophenolic acid morpholinoethyl ester or RS 61443 or RS-61443 or mycophenolate mofetil or Mycophenolate mofetil hydrochloride or MMF).mp.

In clarification, the company confirmed it should read:

(cellcept or mycophenolic acid or morpholinoethyl ester or RS 61443 or RS-61443 or mycophenolate mofetil or Mycophenolate mofetil hydrochloride or MMF).mp.

Novartis have provided the ERG with a list of the 7 unique studies resulting from correcting this error. These studies were double-screened and all 7 studies were excluded. The ERG are content that this point has been dealt with satisfactorily.

- ### Adverse events

The company used their clinical effectiveness search strategy to identify studies reporting adverse events. This strategy worked as the company did not limit their clinical effectiveness searches by study design (i.e. to RCTs using an RCT search filter). The ERG are happy to accept these searches.

4.1.2 Statement of the inclusion/exclusion criteria used in the study selection and comment on whether they were appropriate

As previously mentioned, Novartis presented four systematic literature reviews to identify relevant published and unpublished clinical data. The inclusion/exclusion criteria set out below in Table 3 was used in the search of RCT studies. The eligibility criteria for the NMA studies is similar to the one presented in Table 3. The company reported that none of the search strategies for non-RCTs or adverse events retrieved any relevant studies.

Text in bold presents additional information provided to the ERG by Novartis upon request for clarification.

Table 3. Eligibility criteria used in the search for RCTs

Criteria	***Inclusion criteria***	***Exclusion criteria***	***Justification***
Population	*Adults on maintenance immunosuppressive therapy following allogeneic liver transplantation.*	• *Any non-liver related transplant patients.* • *Pre-liver transplant patients including induction therapy.* • *Non-human.* • *Paediatric.*	*The expected license for everolimus is in combination with reduced dose tacrolimus as immunosuppressive therapy for the prophylaxis of graft rejection after liver transplantation. Initiation of treatment begins 4 weeks after transplantation; therefore the submission addresses maintenance therapy only.*
Intervention/ Comparators	*Everolimus in combination with reduced dose tacrolimus with or without corticosteroids./* ***Azathioprine or mycophenolate mofetil in combination with a calcineurin inhibitor (ciclosporin or tacrolimus) with or without corticosteroids***[2]*.*	• *Sirolimus regimens.* • *Any regimen in combination with an induction treatment.* • *Any study without information on dosages.*	• *The clinical evidence informing the submission relates to everolimus in combination with reduced dose tacrolimus.* • *In clinical practice, corticosteroids are withdrawn or reduced in dose in immunosuppressive regimens depending on patient response in order to reduce the risks of long-term adverse events associated with corticosteroid use. Therefore studies with or without corticosteroids were included to take into account clinical practice.*
Outcomes	• *Patient survival;* • *Graft survival;* • *Time to acute rejection;* • ***Time to recurrence of***	• *Studies that do not focus rejection of the liver as an outcome (efficacy) or HRQL.*	*These include all the outcomes derived from the final scope issued by NICE, except for the following:* • *Graft function / fibrosis*

[2] Provided by Novartis on clarification stage

	hepatocellular carcinoma[2] *;* • *Renal function;* • *Time to end-stage renal disease;* • *Adverse effects of treatment;* • *HRQL.*	• *Studies with only cost and no clinical outcomes.*	• *Expert opinion advised that in clinical experience there was limited evidence to differentiate between interventions with regard to the two outcomes above.*
Study design	• *RCTs of any duration, including cross-over RCTs if data were presented at cross-over.* • *Studies published as abstracts or conference presentations were eligible for the primary analysis of clinical effectiveness if adequate data are provided.*	• *Non-RCT study designs or articles reporting results of RCTs published elsewhere, e.g. reviews, meta-analyses/pooled analyses, editorials, notes, comments or letters.*	• *Only RCTs were considered in line with the objective of this literature search.*

Source: Submission Table 4

Superseded
See Erratum

To note is that for the NMA the company decided to develop a refined criteria as to include any study that included two or more of the following comparators within the study:

1. *Everolimus plus reduced dose tacrolimus with or without a corticosteroid.*
2. *Any combination of MMF and a calcineurin inhibitor (reduced/standard ciclosporin or reduced/standard dose tacrolimus monotherapy) with or without a corticosteroid.*
3. *Any combination of azathioprine and a calcineurin inhibitor (reduced/standard dose ciclosporin or reduced/standard dose tacrolimus monotherapy) with or without a corticosteroid.*
4. *Tacrolimus monotherapy with or without corticosteroid.*

The scope defined the intervention as EVR in combination with TAC and a corticosteroid, however the decision problem addressed in the submission looked at the use of EVR+rTAC with or without corticosteroids. The specification of reduced TAC seems appropriate in theory as the indication of EVR is in combination with a reduced dose of TAC. However, as previously mentioned the reduced TAC dose in H2304 is the equivalent to a standard dose of TAC in UK practice.

After a clarification request from the ERG with regards to the inclusion of ciclosporin in the undertaken literature reviews, the company stated that *all relevant comparators AZA or MMF in combination with a calcineurin inhibitor (ciclosporin or TAC) with or without corticosteroids* were included in all the literature searches. However, after consultation with clinical experts, the company decided to exclude ciclosporin from the economic analysis. This is appropriate as previously explained by the ERG.

After clarification, the company confirmed that mycophenolic mofetil was used in the submission instead of mycophenolic acid, originally defined in the scope. The company claims that there are two presentations of mycophenolic acid (mycophenolate mofetil and mycophenolate sodium) but that mycophenolate sodium is not licensed for use in liver transplantation (only for renal transplantation). Clinical opinion sought by the ERG informed that in the UK most clinicians will prescribe

mycophenolate acid and that even though mycophenolate sodium and mofetil cannot be interchangeably used, the active component is the same and only a dose adjustment is necessary as 500mg mycophenolate mofetil = 360mg of mycophenolate sodium. Therefore, the ERG do not see a valid reason to specify the type of mycophenolate in the inclusion criteria, and the broad mycophenolate acid term could have been used instead.

Novartis clarified that time to recurrence of hepatocellular carcinoma was included in the searches while graft function and graft fibrosis were omitted as per expert opinion.

The submission includes a flow diagram that shows the number of studies identified through the database searches and the number of studies included and excluded at each stage of the review and the reasons for exclusion.

Superseded

Overall, the inclusion criteria seems appropriate to identify all the relevant evidence set out in the NICE scope.

See Erratum

4.1.3 Studies included and excluded

The search strategy identified one RCT, H2304 which studied the intervention of interest (EVR+rTAC with or without corticosteroids) and 33 individual records which were related to this RCT. These can be found in Table 5 (page 44) in the submission.

To note is that the company use different references to quote H2304 throughout the submission as some of them refer to papers, posters or presentations. H2304 is a phase III RCT for which there are 3 main clinical study reports (CSRs), corresponding to a 12-month, 24-month and a 36 months analysis to evaluate the long-term efficacy and safety of concentration-controlled EVR in liver transplant recipients. The main references for these are:

- Hexham et al, 2011 – CSR for the 12-month analysis
- Lopez et al, 2013 - CSR for the 24-month analysis
- Rauer et al, 2014 - CSR for the 36-month analysis

A brief description of the H2304 trial is given in Table 4 below. The company mainly quotes de Simone et al, 2014 as the reference for H2304. This is a study funded by Novartis, looking at the 12-month data from Hexham et al, 2011 CSR.

Table 4. Details of the H2304 Study

Trial name	***Population***	***Interventions***	***Comparator***	***Primary study reference***

H2304	*Adult (18-70 years) de novo liver transplant recipients, who have received a full size liver transplant from a deceased donor; recipient has been initiated on an immunosuppressive regimen containing tacrolimus and corticosteroid (with or without mycophenolate mofetil).*	*1) everolimus (2mg daily dose) + reduced dose tacrolimus (3-5 ng/mL) + corticosteroids* *2) everolimus (2mg daily dose) + tacrolimus elimination (3-5 ng/mL until complete elimination) + corticosteroids*	*Standard-dose tacrolimus (8-12 ng/mL reduced to 6-10 ng/mL at month 4 until end of study) + corticosteroids.*	*(de Simone et al,2014).*

Source: Novartis Submission Table 6

As reported in Table 4, study H2304 was initiated with 3 treatment arms, however, enrolment to the TAC elimination arm was stopped by an independent data monitoring committee (DMC) due to a higher rate of acute rejection and treatment discontinuations in this group. Upon the DMC findings, patients who had not reached day 180 post-randomisation were discontinued from medication and converted to local standard treatment. For patients who were beyond their day 180 post-randomisation there was a choice to either continue in this arm or be converted to a local standard of care. All patients who converted to standard of care were followed up. Irrespective of patients' decision, all patients were to remain in the study for additional safety follow-up through the end of study up to month 24. Following successful completion of month 24, patients had the choice to continue on the assigned regimen for the extension study up to month 36.

As per the DMC recommendations, comparisons on efficacy and safety were mainly made between the EVR+rTAC arm and the TAC control treatment arms.

A brief description of the studies excluded and reasons for exclusion are presented in Table 5 below.

Table 5. Excluded studies containing a regimen of everolimus in combination with tacrolimus

Study	***Everolimus arm***	***Reason for exclusion***	***Reference***
RESCUE	*Everolimus with reduced or eliminated calcineurin inhibitor with/without steroids*	*Both tacrolimus and ciclosporin combined with everolimus and mycophenolic acid (MPA) and azathioprine also added to regimens*	*(De Simone et al., 2009)*
PROTECT	*Everolimus with calcineurin inhibitor*	*Everolimus in combination with calcineurin inhibitor withdrawal, not calcineurin inhibitor reduction*	*(Fischer et al., 2012)* *(Schlitt et al., 2011)*
EPOCAL	*Everolimus with withdrawn calcineurin inhibitor*	*Everolimus with tacrolimus discontinuation after 30 days*	*(Cillo et al., 2013)*
Study not specified	*Everolimus with standard dose tacrolimus*	*Not reduced dose tacrolimus*	*(de Ruvo et al., 2012)*
H2307	*Everolimus with reduced dose tacrolimus*	*Study in progress, no data*	*(Song et al., 2013)*
Data from H2304	*Everolimus with reduced dose tacrolimus*	*Data not reported sufficiently; only the efficacy response relationship and contribution of everolimus to efficacy reported.*	*(Junge et al., 2013)* *(Junge et al., 2013)*

Source: Novartis Submission Table 7

The reasons for excluding some of the documents presented in Table 5 were not very clear to the ERG. Therefore we analysed some of the studies for further details. These are now discussed.

For the Ruvo et al (2012) study, despite the company's claim that it didn't include a reduced dose of TAC, the study reported that immunosuppression blood trough levels were reduced by half at 6 and 12 months in clinically stable patients. However, the study's main focus was on FOXP3+ Treg cells and therefore not suitable for this analysis.

H2307 anticipate results in early 2016, although, recipients were mainly enrolled from Asia and from living donors, so a closer analysis would be required when further data is available.

Junge et al 2013 analysis was primarily based on potency of EVR and TAC and therefore correctly excluded in the submission.

As previously mentioned, it would be useful to compare EVR monotherapy with EVR+rTAC. The EPOCAL (Cillo et al, 2013) and the PROTECT (Fisher et al, 2012 and Schlitt et al, 2011) studies provide insight into this matter. Therefore we give a brief description of these studies:

EPOCAL study

Cillo et al, 2013 performed a phase 2, multicentre, randomized, open label, clinical trial to assess the safety and efficacy of early introduction of EVR in patients undergoing first time liver transplantation. EVR was initiated and TAC weaned in the intervention study group whilst conventional immunosuppression was given in a control arm. The primary endpoint was incidence of acute rejection and graft loss in the first 3 months after liver transplantation. Secondary endpoints were renal function, suspension of TAC and incidence of adverse events.

Study preliminary results showed that no significant differences were found between the two groups. Acute rejection was reported as 17% vs 7% ($p>0.05$) and graft loss was 2% vs 8% ($p>0.05$) in the study and control group respectively. The number of adverse events was also comparable between the two groups with the exception of infections having a significantly higher incidence in the controls (20% vs 1%, $p < 0.01$). It was concluded that the weaning (i.e elimination) of TAC at 1 month in the study group was successful in 43% of cases and EVR seems safe and effective when introduced early after liver transplantation. Renal function at 3 months in the two groups was not significantly different.

Further details regarding conventional immunosuppression and the weaning of TAC in the EPOCAL study were not provided.

PROTECT study

In this 11 month multicentre, open-label study, 4-8 weeks post-liver transplant patients with good renal function were randomised to either continue treatment with CNI (ciclosporin and steroids) or switched to another treatment arm, where they would initiate treatment with EVR and steroids.

Results showed that the mean difference in eGFR calculated by the four-variable modification of diet in renal disease (MDRD4) at month 11 between the EVR group and the CNI group was 7.778 ml/min (P = 0.0209, ITT population). Furthermore, Schlitt et al 2011 stated *conversion to an EVR-based regimen showed comparable efficacy at month 11 evaluations for incidence of biopsy proven acute rejection (BPAR), graft loss or death*. These results demonstrate that in liver transplant recipients with good renal function 4-8 weeks post-transplant, conversion to EVR (mostly monotherapy) can be achieved without compromising efficacy and may lead to an improvement in renal function during the first post-operative year. However, further study information would be required for analysis and validation.

Both the EPOCAL and the PROTECT study seem to show that EVR monotherapy is as effective as other treatments (containing TAC or not) in maintaining a functioning liver graft, whilst preserving renal function.

Comparators - NMA

Since there was no evidence which allowed a direct comparison between EVR+rTAC and the comparator treatments, Novartis undertook a NMA in order to estimate the relative effectiveness between all treatments.

After a full-text review, a set of 61 records and 22 RCTs were included for review. Following an assessment of the comparators in the 22 RCTs, the following comparator arms were found to enable a network of studies:

- Everolimus plus reduced dose tacrolimus (EVR+rTAC)
- Mycophenolate mofetil plus standard dose tacrolimus (MMF+TAC)
- Mycophenolate mofetil plus reduced dose tacrolimus (MMF+rTAC)
- Mycophenolate mofetil plus standard dose ciclosporin (MMF+CIC)
- Azathioprine plus standard dose tacrolimus (AZA+TAC)
- Azathioprine plus standard dose ciclosporin (AZA+CIC)
- Standard dose tacrolimus (TAC).

When referring to the 22 RCT studies identified through their searches, the company stated that *although it was unclear if randomisation or concealment of treatment allocation was appropriate in most studies, no studies were excluded due to risk of bias*.

Due to a wide range of studies considered for the NMA and time constraints, the ERG approach was to focus on:

- The studies included in the final NMA
- The studies used to derive the parameters included in the economic model (tBPAR and renal functioning)

Nevertheless, the ERG are still not clear which studies have been included in the final NMA analysis for the biopsy proven acute rejection (tBPAR) outcome, due to lack of clarity and transparency in the submission. This is further explored in this section.

4.1.4 Details of any relevant studies that were not included in the submission

The ERG considers that all studies relevant to assessing EVR in combination with reduced TAC against MMF+TAC and AZA+TAC with respect to the scope of this submission were included.

4.1.5 Description and critique of company approach to validity assessment

A critique of the company's approach to the analysis of relevant RCTs is provided below alongside comments made by the ERG on its appropriateness.

Table 6. Clinical appraisal of relevant RCTs

Assessment question	***Novartis response***	***Grade (yes/no/not clear/N/A)***	**ERG comments**
Was randomisation carried out appropriately?	*Randomisation was carried out using an interactive voice response system.*	*Yes.*	The company's submission reports that patients were stratified according to pre-transplant hepatitis C status and quartiles of renal function in order to balance the risk factors for graft and patient survival and then randomised in a 1:1:1 ratio. An interactive list was produced by the interactive voice response system (IVRS). The ERG considers this method appropriate.
Was the concealment of treatment allocation adequate?	*Yes. The randomisation numbers were generated using the following procedure to ensure that treatment assignment was unbiased and concealed from patients and investigator staff. A patient randomisation list was produced by the IVRS provider using a validated system that automated the random assignment of the patient numbers to randomisation numbers.*	*Yes*	In addition to using an IVRS, the company's submission also reported that the randomisation scheme for patients was reviewed and approved by a member of the Biostatistics Quality Assurance Group. The ERG considers this method appropriate.

Were the groups similar at the outset of the study in terms of prognostic factors, for example, severity of disease?	*Baseline characteristics appear similar in all treatment arms.*	*Yes*	On clarification, the company stated that there was a mistake with the reporting of HCV and HCC rates at baseline in treatment arms. Once this was clarified, the ERG was satisfied that the baseline groups were similar.
Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?	*No. Open-label trial, it is unclear whether blinding would be suitable for this trial.*	*No*	Most of the clinical studies looking at transplantation patients included in the submission were either open-label or did not report blinding. However, Weisner et al 2001, reported a double-blinded trial. The company also reported that the investigator, pharmacist and patient were aware of the treatment. However, the treatment assignment was withheld from the local pathologist interpreting the biopsies. It is not clear the impact this had on study outcomes.
Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	*Yes. The tacrolimus Elimination group was terminated prematurely based on a recommendation of the Data Monitoring Committee due to a higher incidence of acute rejection versus the other cohorts.*	*Yes*	Study drop-out was greater in the EVR+rTAC arm than in the TAC control arm at 12 months (66 vs 54) and 24 months (104 vs 79) respectively. Explanations provided in the CSR are unclear.
Is there any evidence to suggest that the authors measured more outcomes than they reported?	*No. All discussed outcomes appear to have been reported.*	*No*	The ERG find the wording in the company reply curious. Trial H2304 was a Novartis study, therefore the company should have full knowledge of the measurement and reporting of study outcomes.

Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	*All randomised patients were included in the intent-to-treat population. This was appropriate but the methods used to account for missing data are unclear.*	*Yes*	Patients with missing data at the 12 month analysis were censored at the last day known to be free of event (tBPAR, graft loss or death). The submission doesn't provide information on the methodology used to deal with this. However, the CSR (Hexham et al, 2011) reports some imputation methods used for renal outcomes. Again, the ERG find it strange that the company are not clear about the methods used to account for missing data given that H2304 is a Novartis trial.

Response source: Novartis Submission Table 98

Novartis have covered the elements used in the critical appraisal of RCTs according to the Centre for Reviews and Dissemination Systematic Reviews checklist (2008).

Additionally, the ERG note that a limited number of UK patients were involved in the trial. The study was conducted predominantly with patients in the US, [redacted].

Superseded
See Erratum

Novartis assessed the NMA studies for their validity (Table 16 in the submission). Even though they follow the template suggested by NICE to assess the NMA studies, the ERG do not find that it provides very useful information as all the questions were answered with a Yes/No/Unclear reply. So for example, one of the criteria used by Novartis to assess the validity of the trials is "Were the groups similar at the outset of the study in terms of prognostic factors?". If the company considered this to be a "Yes", there is no way for the ERG to assess this answer as the submission does not state the criteria used to assess similarity in terms of prognostic factors.

Also the proportion of answers answered with "unclear" was considerably large. For example, for the first NMA question, 'Was randomisation carried out appropriately?' 68% of the answers were answered "unclear". Overall for all questions answered, 40% were answered "Unclear".

Studies were also grouped into categories within the critical appraisal section, so it was difficult to assess how each study had been individually appraised.

4.1.6 Description and critique of company's outcome selection

- H2304 study

Primary Efficacy Endpoint

The primary efficacy end-point in H2304 was a composite of 3 outcomes defined in the scope. These are the efficacy failure rate of treated biopsy proven acute rejection (tBPAR), graft loss or graft death at 12 months post transplantation (excluding any events before randomisation).

tBPAR was determined as acute rejection with a locally confirmed rejection activity index (RAI) ≥ 3 according to the Banff 1997 criteria when treated with anti-rejection therapy. The Banff RAI includes 3 components scored from 0 to 3: venous endothelial inflammation (E), bile duct damage (B), and portal inflammation (P). The scores are combined to an overall score (the RAI).

Although the selection of the primary outcomes measures seems to be appropriate to evaluate the primary efficacy end-points in H2304, clinical opinion sought by the ERG explained that given the rapid evolution of transplantation medicine, the efficacy endpoints used in the trial might not be the most appropriate ones as explained in our Section 3.3.

Secondary Efficacy Endpoint

The key secondary outcome was change in renal function over time from randomisation to 12 months post-transplantation measured by the estimated glomerular filtration rate (eGFR). The four-variable modification of diet in renal disease (MDRD4) formula was used for eGFR measurement.

Other secondary outcomes included:

- *Incidence of a composite of treated biopsy proven acute rejection (BPAR), graft loss, death or loss to follow-up.*
- *Incidence of each component of the primary composite efficacy endpoint.*
- *Incidence of a composite of death (D) or graft loss (GL).*
- *Treated BPAR endpoint defined by: (1) incidence, (2) time to event, (3) severity, (4), diagnosis leading to transplantation.*
- *Acute rejection by: (1) incidence, (2) time to event, (3) severity.*

Renal function-related secondary outcomes were also included:

- *Evolution of post-randomization renal function over time assessed by the change in eGFR (MDRD-4) from baseline to Months 12 and 24.*
- *Renal function by eGFR using various methods (MDRD-4/6, Nankivell, Cockcroft-Gault, Chronic Kidney Disease Epidemiology Collaboration [CKD EPI] and Hoek formulae).*
- *Incidence of patients experiencing a decline in eGFR of < 10, 10-<15, 15-<20, 20-<25, and ≥25 mL/min/1.73m2 from Screening, Week 2 post transplantation and randomization to Months 6, 12 and 24.*
- *Serum creatinine at various time points.*
- *Evolution of renal function by chronic kidney disease (CKD) strata.*
- *Renal function and change in eGFR from screening, Week 2 post transplantation and randomisation to Months 6 and 12 eGFR in the following subgroups:*
    - *Age (<60 and ≥60 years), gender, race, region, renal function strata (< 30, 30-<45, 45-<60, ≥60 mL/min/1.73m2, below/above 45 mL/min/1.73m2, below/above 60 mL/min/1.73m2), HCV status, Model for End-Stage Liver Disease (MELD)*

score categories (≤14, 15-19, 20-24, 25-29, ≥30), and diagnosis leading to transplantation.

- *Urinary protein/creatinine ratio at various time points.*
- *Incidence of proteinuria of 0.5- <1.0 g/day, 1.0-<3.0 g/day and ≥3.0 g/day at various time points.*
- *Incidence of and time to renal replacement therapy.*

Safety related outcomes consisting of:

- *Incidence of AEs/infections/serious adverse events.*
- *Incidence of treatment-related side effects and other AEs of interest, including incidence of new onset diabetes mellitus (NODM), evolution of metabolic parameters as subdivisions of serum/plasma lipid panel, neurotoxicity and hypertension.*
- *Incidence and reason (e.g. AE) of premature discontinuation of study medication and premature discontinuation from the study.*
- *Incidence and reason (e.g. AE) of interruption and dose adjustment of study medication.*
- *Incidence of patients in the tacrolimus elimination arm that needed CNI treatment.*

Hepatitis C virus (HCV) and Hepatocellular Carcinoma (HCC) related outcomes were also reported:

- *Evolution of HCV and HCV related fibrosis*
- *HCV viral load (HCV-RNA) levels overall and by genotype.*
- *Rates of progression of HCV related allograft fibrosis.*
- *Incidence of and response to HCV antiviral treatment.*
- *Recurrence of HCC and incidence of de novo HCC malignancies.*
- *Rate of recurrence of HCC at 12 and 24 months post-transplantation in patients with a diagnosis of HCC at the time of liver transplantation adjusting for various risk factors, such as number of tumor nodules, total tumor diameter, alpha fetoprotein level, etc.*
- *Incidence of de novo HCC malignancies*

Source: Submission section 10.16 Appendix 16

The selection of outcomes appears reasonable to provide a sensible range of dimensions to assess the clinical effectiveness of EVR and its renal sparing effect.

Unfortunately there was no health questionnaire to directly capture patients' Health related Quality of Life (HRQL) in the H2304 trial.

- NMA

After selecting the relevant clinical studies, Novartis conducted a feasibility assessment of the NMA, which was based on 16 clinical endpoints extracted from the studies. Table 7 below highlights the 16 efficacy and safety endpoints chosen.

Table 7.Description of the 16 endpoints considered for the NMA

Endpoints	*Definitions*

Efficacy	
Survival at 12 months	• *The event is defined as death from any causes.* • *The Kaplan-Meier method should be used to calculate the event free probabilities.*
Survival at 24 months	
Graft survival at 12 months	• *The event is defined as death from any causes or graft loss.* • *The Kaplan-Meier method should be used to calculate the event free probabilities.*
Graft survival at 24 months	
Acute rejection at 3 months	• *The event is defined as Biopsy Proven Acute Rejection (BPAR).* • *The Kaplan-Meir method should be used to calculate the event free probabilities.*
Acute rejection at 6 months	
Acute rejection at 12 months	
Acute rejection at 24 months	
Renal function at 12 months	• *The renal function could be measured by the estimated Glomerular Filtration Rate (eGFR) using any formula, e.g. MDRD4, or by the estimated Creatinine Clearance Rate (eCCR) using the Cockcroft–Gault formula.* • *For any measurement tool, Change from Baseline (CFB) should be used.*
Renal function at 24 months	
Safety	
Withdrawals due to Adverse Events (AEs)	• *Number of withdrawals due to AEs.*
Hypertension	• *Number with hypertension.*
Diabetes	• *Number of general diabetes or new-onset or de novo diabetes or new-onset and worsening of pre-existing diabetes.*
Infections	• *Number of any infections.* • *If multiple kinds of infections are reported, the highest number was extracted.*
Hepato-Cellular Carcinoma (HCC)	• *Number of recurrent HCC.*
Hepatitis C Virus (HCV)	• *Number of new-onset or recurrent HCV.*

Source: Novartis Submission Table 17

Upon clarification stage, Novartis stated that the selected endpoints included relevant key efficacy outcomes with regards to immunosuppressive treatment, including acute rejection, overall survival, graft survival and renal function. Safety outcomes were said to have been selected based on previous clinical trials in hepatic immunosuppression and confirmed to be of relevance via clinical expert advice. Recurrence of hepato-cellular carcinoma and new onset / recurrence of HCV infection were regarded as important factors in hepatic transplantation, with a large proportion of hepatic transplants being due to progression of chronic HCV infection or HCC.

The two research questions considered by the company for the assessment of feasibility of the NMA were:

1. Is there a network of interlinked RCTs to allow the comparisons of interest to be estimated?

2. Are there any differences in study and patient characteristics across studies that could affect the estimates of the effectiveness of the intervention against the relevant comparators?

Results of feasibility assessment

The evidence network was deemed feasible for 13 out of the 16 outcomes reported in Table 7 above. The company reported that no network was produced for HCC recurrence at 12 months, renal function and tBPAR at 24 months.

HCC recurrence was excluded as only 2 RCTs, H2304 and Boudjema et al, 2011 reported data for this outcome and the number of events was reduced, which could have caused convergence issues on running the NMA analysis.

For tBPAR at 24 months, Novartis mention 3 studies, namely, H2304, Jain et al 2001 and Dmitrewski et al 1994 which reported this outcome. The company claims that this would have produced a reduced set of comparators which wouldn't have contributed to the decision problem. However, the ERG are not convinced by this statement as a network could have been built for all the relevant comparators EVR+rTAC, MMF+TAC and AZA+CIC all linked via a standard TAC arm.

Mean eGFR at 24 months was only reported in H2304.

AEs were only analysed at 12 months. It is not clear why the company didn't consider longer term data for AEs, for example 24 month data.

The answer to the second feasibility question identified treatment modifiers such as exclusion of concomitant steroids in the first 3 months, HCC, cold ischaemic time, donor age, hypertension and diabetes. Novartis reported that more patients in H2304 had diabetes and hypertension than in any other reported study.

The ERG found variation in potential treatment modifiers reported throughout the studies. For example, Samuel et al 1998 reported 'total ischaemic time', instead of referring to 'cold ischaemic time'. Furthermore, Jain et al 2001, reported donor age, however, no other potential treatment modifiers were included such as cold ischaemic time, HCC, diabetes or hypertension.

After discussion with a clinical expert, it was also established that a '*Donor Risk Index*' would have provided a useful and unified measure for potential treatment modifiers.

The standard TAC arm of H2304 was also reported as having better overall survival, graft survival and free from tBPAR probabilities at 12 and 24 months than the standard TAC arms of other reported studies. The company could not explain this and considered H2304 to be comparable to the rest of the studies, thus decided to attribute this discrepancy to chance. The ERG found several discrepancies between H2304 and the other NMA studies. The main one is the trough levels of TAC observed in what it was considered as "standard" TAC arm by the company.

The ERG found that many of the studies used in the NMA analysis had significantly different dosing levels to UK clinical practice. Within the UK and according to the Prograf SPC protocol, standard TAC trough levels are 5-20ng/ml and reduced TAC trough levels are ≤5ng/ml. Many studies maintained TAC trough levels above 5ng/ml in their reduced TAC arms such as Neuberger et al 2009, whilst other studies such as Porayko et al, 1994 maintained TAC trough levels below 5ng/ml in their

standard TAC arms. Consequently, no consistency was seen across studies with respect to drug levels. This is further explored throughout the report.

The company reported heterogeneity and variation across studies which provided challenges to building a feasible network. The main variation across studies were identified to be:

- *Lack of reporting in studies of the characteristics that could be potential treatment effect modifiers to enable assessment to be made across the set of studies.*
- *Variation in the definition of the tacrolimus and ciclosporin arms with respect to dosage across the studies (definitions between standard dose and reduced dose).*
- *Variations in the definitions for the outcomes included across studies, for renal function, diabetes and infections.*
- *Variation in the duration of use of corticosteroid therapy in combination with the interventions of interest.*

4.1.7 Description and critique of the statistical approach used

- H2304 study (12 months)

tBPAR, graft loss or death

Statistical analysis performed in the H2304 trial was designed to show the non-inferiority of EVR+rTAC compared to control in terms of the composite efficacy failure of tBPAR, graft loss, or death (tBPAR/GL/D) at 12 months by rejecting the null hypothesis that the control is superior to treatment against the alternative that EVR+rTAC is not inferior to the control:

H_O: $\pi T - \pi C \geq 0.12$

H_A: $\pi T - \pi C < 0.12$

where πT and πC were the true proportions of composite efficacy failure of tBPAR/GL/D at 12 months post-transplant for the respective treatment groups. A pre-determined non-inferiority margin of 12% was used for the analysis.

The primary efficacy failure proportions per treatment arm were estimated using the Kaplan-Meier (KM) product-limit formula for the ITT population.

The non-inferiority test for the proportion of composite efficacy failure was performed using the two-sided 97.5% confidence intervals, based on Z-test, for the between-treatment difference of the proportions of efficacy failure ($\pi T - \pi C$). If the upper bound of the interval was below 12%, the EVR+rTAC regimen was reported as non-inferior to control on the composite failure outcome.

In non-inferiority trials, the choice of the non-inferiority margin is crucial. Novartis do not explain their decision in their submission. However the Hexham et al, 2011 CSR reports that the non-inferiority margin should be such that it is less than the upper limit of the 95% CI of the control effect between the control arm (TAC plus corticosteroids) and the putative placebo (reduced TAC plus corticosteroids) but that based on Medline search results, there were no published liver transplant trials directly comparing these and no published trials using the putative placebo as a treatment arm where the intervention was begun 30 days after liver transplantation. Due to the lack of historical data, the choice of 12% non-inferiority margin was based on the 10% non-inferiority margin chosen for the graft loss or death endpoint in a study using an AZA + steroids regimen (historical control) as the putative placebo. Since the endpoint in the study only included graft loss or death, the inclusion of tBPAR as a clinical endpoint would likely lead to a higher incidence of this composite endpoint in the control arm of H2304.

On a relative scale, the risk ratio for the maximum allowed rate of the original primary endpoint of graft loss or death was 2.25 (18% vs. 8%). With the assumed background rate of 20% to 24% for the new endpoint, the risk ratio was <1.6 with an NI margin of 12%. Therefore, an increase of 2% over the absolute NI margin for the graft loss or death endpoint was judged to be appropriate for the composite efficacy failure endpoint of tBPAR, graft loss, or death (Hexham et al, 2012). Further information should have been provided to fully justify the appropriateness of a 2% increase over the original 10% margin.

A sample size of 242 patients was used alongside a power value of 80% at the one-sided 0.0125 level to demonstrate inferiority of the EVR+rTAC group versus the TAC control arm in the proportion of patients with tBPAR, graft loss or death.

For secondary outcomes including, among other measures, acute rejections, treatment comparisons between the EVR+rTAC regimen and the control were made over the ITT population using the Pearson Chi-square test. Exact tests were used for events such as, graft loss or death, with low event rates (<5%).

Renal sparing effect

The evolution of renal function was also evaluated by a non-inferiority test to show non-inferior renal function of the EVR+rTAC regimen to the control, for a mean change in eGFR from randomisation to month 12. This assumed a non-inferiority margin of the difference in eGFR of 6 mL/min/1.73m2 from randomization to 12 months post-transplant. The ERG couldn't find a justification for the non-inferiority margin used.

If the null hypothesis of inferiority in terms of renal function was rejected, the EVR+rTAC regimen was further tested against the control for superiority.

With N=217 patients per arm, the study had at least 90% power to show that EVR+rTAC was not inferior to the control arm on mean change in eGFR at 12 months post-transplantation at the two-sided 0.025 level, assuming the non-inferiority margin for the difference of change in eGFR was -6.0 mL/min/1.73m2, the standard deviation was 20 mL/min/1.73m2, and the correlation coefficient with randomization eGFR was 0.5.

Censoring

For patients who had prematurely discontinued the study drug, primary and secondary efficacy data was missing at the 12 and 24 month analyses points. Patients with missing primary and secondary efficacy evaluation for the 12 month analysis were censored at the latest day known to be free of the event. More information of censoring would have been useful.

The overall approach to the statistical analysis of H2304 was generally sound.

- Sub-group analysis

A further sub-group analysis was reported for the primary efficacy endpoint tBPAR, graft loss and death according to region (North America, Europe, Rest of the World), age group (< 65 versus ≥ 65 years), gender, race (caucasian, black, asian and all others), renal function strata at randomisation(<30, 30-<55, 55-<70,70-<90 and ≥90) HCV status (negative or positive), recipient MELD score categories ≤14, 15-19, 20-24, 25-29 and ≥30) and types of end stage disease leading to transplantation. The efficacy failure event time distributions were presented using Kaplan-Meier estimates, and comparisons between treatment groups were made using the proportional hazards model adjusting for pre-transplantation HCV status (95% CI).

eGFR and change in eGFR from screening to week 2 post-transplantation and from randomisation to Month 6 and Month 12 are also reported in the following subgroups based on *the intention-to-treat (ITT) population: age group, gender, race, renal function strata based on eGFR (MDRD-4) at randomisation (strata were: (a) < 30, 30 - 44, 45 - 59, ≥ 60 mL/min/1.73m2; (b) below/above 45 mL/min/1.73m2; and (c) below/above 60 mL/min/1.73m2), HCV status, MELD score categories (≤ 14, 15 - 19, 20 - 24, 25 - 29, ≥ 30), and diagnosis (end stage disease) leading to transplantation.*

Subgroups were pre-planned and chosen on the basis that they would be expected to have differential responses to immunosuppressive therapy in terms of efficacy and adverse event incidence.

- H2304 study (24 months)

The 24 month statistical analysis was similar to the one undertaken for the 12 month analysis.

- H2304 study (36 months)

The 36 analysis considered the evaluation of long term safety and efficacy of EVR. The main focus of the statistical analysis was the evaluation of renal function, maintenance of efficacy, adverse events, infections and laboratory values. Composite efficacy endpoints for tBPAR, graft loss and death from the extension baseline at month 24 to month 36 were summarised using frequency counts and percentages by treatment group.

The sample size reported for the extension study consisted in patients enrolled into the extension period. 282 patients have been reported to enter the extension study from the initial 719 patients entering the main trial.

Descriptive statistics were used to analyse renal function. Changes in eGFR from randomisation and from the start of month 24 were presented.

No subgroup analysis was reported.

- NMA

The ERG found a significant limitation in the NMA presented by Novartis. Even though WinBUGS codes were included in the submission, these did not provide any relevant information. Whilst the codes presented reflect a standard NMA approach, these seem to have been either taken from other submission or from a theoretical exercise, as the data included in the codes do not relate to the submission data in any way. This poses a major limitation as the ERG could not verify which data were used for the analysis of specific outcomes. This represents a major concern as the ERG are not clear which studies have been included in the NMA analysis for the tBPAR outcome, due to lack of clarity and transparency in the submission.

Superseded
See Erratum

The NMA was conducted using a Bayesian framework. The WinBUGS/OpenBUGS software package was used to estimate the parameters of the different NMA models using a Monte Carlo Markov Chain (MCMC). Likelihood and link functions were defined for the different types of outcome data. Non-informative prior distributions were used for model parameters,

A non-informative prior distribution was used for the model parameters. Nuisance and treatment effect parameters followed a normal distribution, μ_{jb}~normal (0, 10,000) and d_{Ak}~normal (0, 10,000) respectively. The heterogeneity parameter was based on a uniform distribution, σ ~ Uniform (0, 5).

The values attributed to the normal distributions were in line with what is suggested in NICE DSU guidance (Dias et al, 2011), however the between-trial variance parameter, σ ~ Uniform (0,5) presents an extremely high upper limit. The ERG requested clarification as to why the upper limit was attributed the value of 5, however Novartis have not clarified this. The fact that this parameter has a high value might be leading to some of the considerably large credible intervals reported for the NMA results.

Both random and fixed effects models were ran with the latter being preferred for most outcome measures. The company presented the Deviance Information Criterion (DIC) and upon the ERG request, the posterior mean of the deviance was also presented to ensure that the selected model's overall fit was adequate. These are presented in Table 8.

In NMAs, convergence of results is normally assessed by using the Brooks-Gelman-Rubin diagnostic in WinBUGS and the Monte Carlo error, which reflects both the number of simulations and the degree of autocorrelation. However Novartis only reported an initial series of iterations which were discarded as 'burn-in' and inferences based on additional iterations. Therefore there was not enough information available to reach conclusions regarding convergence of results'

Table 8: Model fit statistics

Outcome Measure	Fixed Effects	Random Effects	Model selected

	DIC	Dbar	pD	DIC	Dbar	pD	
Overall Survival 12 months	133.3	115.3	17.9	134.8	115.3	19.5	FE
Overall Survival 24 months	83.2	71.4	11.8	85.0	71.8	58.6	FE
Graft survival 12 months	127.8	97.8	15.9	129.6	98.5	16.9	FE
Graft survival 24 months	86.3	74.3	11.9	88.2	75.1	13.1	FE
tBPAR 3 months	138.3	121.4	16.9	139.9	120.9	18.9	FE*
tBPAR 6 months	79.9	68.9	11.0	79.5	67.4	12.0	FE*
tBPAR 12 months	119.7	105.6	14.1	119.3	102.5	16.9	FE*
Renal function (eGFR) 12 months	-0.8	-3.8	3.0	0.3	-4.2	4.4	FE
Withdrawals due to AEs 12 months	68.6	57.7	10.9	68.0	56.7	11.4	FE
Hypertension 12 months	75	65.0	10.1	76.9	65.8	11.1	FE
New onset diabetes 12 months	72.1	62.1	10.0	74.0	62.9	11.0	FE
Infections 12 months	117.0	102.0	15.0	119.1	102.7	16.4	FE
HCV recurrence 12 months	71.5	61.6	10.0	73.1	61.9	11.2	FE

Source: Adapted from Novartis clarification document. ; DIC: Diagnostic Information Criteria, Dbar; posterior mean deviance; pD: effective number of parameters (model complexity=DIC-Dbar)

*It is not clear at all in the submission if FE or RE models were chosen for some of the outcomes. The ERG have based their decision that the FE models were used by Novartis based on the sentence provided in the clarification document "*The fixed effects models were chosen for use in the de novo cost-effectiveness model for the outcomes used in the model".*

Even though the differences between DIC estimates are not very significant (note that smaller values indicate better fit), the ERG agree that the fixed effects model is the best model to use in this analysis, across all outcomes. The main reason for this is that the random effects model adds heterogeneity parameters to the estimation and since the prior distribution for these presents such a high upper limit (5) hence such small precision, the random effects models presented unreasonably wide credible intervals, like for example between 0 and 137200.

4.1.8 Summary statement

Following responses to the ERG's questions for clarification in relation to the effectiveness searches, we are content that the searches presented in this submission are suitable for the task.

The methodology used to assess the quality of the included RCT was adequate, whilst the methodology used to assess the quality and the comparability of NMA studies could have been more informative.

The ERG found a significant limitation in the NMA presented by Novartis as the data included in the WinBUGS codes do not relate to the submission data in any way. This poses a major limitation as the ERG could not verify which data were used for the analysis of specific outcomes. This represents a concern as the ERG are not clear which studies have been included in the NMA analysis for the tBPAR outcome, due to lack of clarity and transparency in the submission.

The ERG consider that the evidence submitted generally reflects the decision problem outlined in the final scope of the submission.

4.2 Summary of submitted evidence

4.2.1 Summary of results

- H2304 study

Efficacy results for hepatic outcomes

The company present H2304 efficacy outcomes for months 12 (Hexham et al, 2011) and 24 (Lopez et al, 2012) on Tables 9 and 10 of their submission. However the ERG found inconstancies in the reporting of results. For example, all the results presented in Table 9 of the company submission report a 95% CI. Nonetheless, this is not the case for all the outcomes as for instance, the main composite efficacy outcome results were obtained through the non-inferiority analysis with a 97.5% CI.

Therefore, the ERG used the CSR documents to report the main efficacy and safety results. Also to note is that Novartis decided not to present the 36 months results (Rauer et al, 2014). Upon the ERG's request for clarification the company explained that these data were very limited. Nevertheless the ERG present the 36 months data in this section.

Of note is that all tables in this section present the results for the TAC elimination arm. Even though these results were not the main focus of the submission the ERG analyses some of these as per reasons explained above.

Table 9, Table 10 and Table 11 below present the comparison between treatment groups for KM-derived incidence rates of primary efficacy outcome (tBPAR, graft loss or death), the KM incidence rates of graft loss or death and the incidence rates of secondary outcomes, respectively. The tables report 12 month analysis based on Hexham et al, 2011.

Finally, Table 12, Table 13 and Table 14 present the same analysis for the 24 months data (Lopez et al, 2012).

Table 9. KM incidence rates of primary efficacy endpoint (ITT population at 12 months)

Statistic	EVR+Reduced TAC N=245	TAC Elimination N=231	TAC Control N=243
Number of composite efficacy failure (tBPAR, graft loss or death) from randomization till Month 12	16	45	23
KM estimate of incidence rate of composite efficacy failure (tBPAR, graft loss or death) at Month 12	6.7%	24.2%	9.7%
Difference in KM estimates (vs. Control)	-3.0%	14.5%	
97.5% CI for difference	(-8.7%, 2.6%)		
P-value of Z-test for (Reduced TAC - Control = 0) (No Difference Test)	[redacted]		
P-value* of Z-test for (Reduced TAC - Control ≥ 0.12) (Non-inferiority Test)	<0.001		

1. tBPAR = treated biopsy proven acute rejection. Local laboratory biopsy results are used to define tBPAR.
2. *Z-test p-value for non-inferiority test (non-inferiority margin = 12%) is for one-sided test and was compared to 0.0125 significance level.
3. In Kaplan-Meier estimate, the censoring day for patients without event is the last contact day.
Source: Table 14.2-1.1a.

Table 10. KM incidence rate of graft loss or death (ITT population at 12 months)

Statistic	EVR+Reduced TAC N=245	TAC Elimination N=231	TAC Control N=243
Number of patients with graft loss or death from randomization till Month 12	12	10	7
KM estimate of incidence rate of patients with graft loss of death at Month 12	[redacted]	[redacted]	[redacted]
Difference in KM estimates (vs. Control)	[redacted]	[redacted]	
97.5% CI for difference	[redacted]		
P-value of Z-test for (vs. Control = 0) (No Difference Test)	[redacted]		
P-value of Z-test for (vs. Control ≥ 0.10) (Non-inferiority Test)	[redacted]		

Source: Table 14.2-2.1e.

Table 11. Incidence rates of secondary efficacy endpoints (ITT population at 12 months)

				EVR+Reduced TAC vs. TAC Control	
Efficacy endpoints	EVR+Reduced TAC N=245 n (%)	TAC Elimination N=231 n (%)	TAC Control N=243 n (%)	Risk Diff. (95% CI)	P-value
Graft loss*	6 (2.4)	5 (2.2)	3 (1.2)	1.2% (-7.8, 10.2)	0.5038
Death*	9 (3.7)	8 (3.5)	6 (2.5)	1.2% (-7.8, 10.1)	0.6015
AR	9 (3.7)	46 (19.9)	26 (10.7)	-7.0% (-11.6, -2.5)	0.0026
[redacted]					
BPAR	10 (4.1)	46 (19.9)	26 (10.7)	-6.6% (-11.2, -2.0)	0.0052
tBPAR	7 (2.9)	38 (16.5)	17 (7.0)	-4.1% (-8.0, -0.3)	0.0345
[redacted]					

1. AR = Acute rejection; BPAR = biopsy proven acute rejection; tBPAR = treated biopsy proven acute rejection. Local laboratory biopsy results are used to define BPAR and tBPAR.
2. * = exact confidence interval and two-sided Fisher exact test used for that variable. For others, asymptotic confidence interval and Pearson Chi-square test are used.
3. All p-values are for two-sided test and were compared to 0.05 significance level.
Source: Table 14.2-2.1a

Table 9 shows that the difference in KM estimates between EVR+rTAC and TAC was 3% in favour of EVR+rTAC, CI (-8.7% to 2.6%). This means that the composite endpoint (tBPAR, graft loss or death) had fewer events observed in the trial in the EVR+rTAC arm than in the TAC arm of the trial. As the upper limit of confidence interval is below 12% and the p-value is reported to be <0.001, EVR+rTAC is said to be statistically non-inferior to TAC in terms of the composite efficacy outcome.

Table 10 reports the statistical comparison between treatment arms for KM incidence rates of graft loss or death. For this analysis, a non-inferiority margin of 10% was used as explained in the previous section. [redacted]

It seems that the composite primary endpoint of the trial (tBPAR, graft loss and death) combines the effectiveness of two outcomes which work in different directions. Whilst graft loss or death seem to favour the TAC arm of the trial, the addition of tBPAR to the composite endpoint favours the EVR+rTAC arm. This casts doubt on the appropriateness of the clinical endpoints as previously discussed.

Table 11 presents the incidence rates for secondary outcomes across treatment arms. The statistical analysis undertaken on these outcomes produced qualitatively different results from those for the primary outcomes. It can be observed that graft loss and death were not statistically different across the treatment arms, with p-values>0.05 and that statistically significant fewer episodes of rejection were found in the EVR+rTAC compared with the TAC arm under all outcome measures (AR, tAR, BPAR and tBPAR). This reinforces the point aforementioned about the inclusion of rejection episodes in the composite primary outcome.

Table 12. KM incidence rates of primary efficacy endpoint (ITT population at 24 months)

Statistic	EVR+Reduced TAC N=245	TAC Elimination N=231	TAC Control N=243
Number of patients with composite efficacy failure (tBPAR, graft loss or death) from randomization to Month 24	24	55	29
KM estimate of incidence rate of composite efficacy failure (tBPAR, graft loss or death) at Month 24	10.3%	26.0%	12.5%
Difference in KM estimates (vs. Control)	-2.2%	13.5%	
97.5% CI for difference	(-8.8%, 4.4%)		
P-value of Z-test for (EVR+Reduced TAC - Control = 0) (No Difference Test)	0.452		
P-value of Z-test for (EVR+Reduced TAC - Control >= 0.12) (Non-inferiority Test)	[redacted]		

tBPAR = treated biopsy proven acute rejection. Local biopsy pathology results are used to define tBPAR.
One-sided Z-test p-value for non-inferiority test (non-inferiority margin = 12%) is presented.
In Kaplan-Meier estimate, the censoring day for patients without event is the last contact day.
Source: Table 14.2-1.1a

Table 13. KM incidence rates of graft loss or death (ITT population at 24 months)

Statistic	EVR+Reduced TAC N=245	TAC Elimination N=231	TAC Control N=243
Number of patients with graft loss or death from randomization to Month 24	17	18	14
KM estimate of incidence rate of patients with graft loss of death to Month 24	7.3%	8.6%	6.2%
Difference in KM estimates (vs. Control)	1.1%	2.4%	
97.5% CI for difference	(-4.2%, 6.4%)		
P-value of Z-test for (vs. Control = 0) (No Difference Test)	0.638		
P-value of Z-test for (vs. Control ≥ 0.10) (Non-inferiority Test)	[redacted]		

1. P-values and 97.5% CI are obtained using Kaplan-Meier (KM) probability estimates of composite efficacy failure rates and standard error derived based on Greenwood's formula.
2. One-sided Z-test p-value for non-inferiority test (non-inferiority margin = 10%) is presented.
3. In Kaplan-Meier estimate, the censoring day for patients without event is the last contact day.

Source: Table 14.2-2.1a

Table 14. Incidence rates of secondary efficacy endpoints (ITT population at 24 months)

				EVR+Reduced TAC vs. Control		
Efficacy variable	EVR+Reduced TAC N=245 n (KM %)	TAC Elimination N=231 n (KM %)	TAC Control N=243 n (KM %)	Diff Risk (KM%)	97.5% CI	P-value
[redacted]						
Graft loss or death	17 (7.3)	18 (8.6)	14 (6.2)	1.1	(-4.2, 6.4)	0.6378
Graft loss	9 (3.9)	6 (2.8)	7 (3.2)	0.8	(-3.2, 4.7)	0.6605
Death	12 (5.2)	15 (7.3)	10 (4.4)	0.8	(-3.7, 5.2)	0.7012
[redacted]						
BPAR	14 (6.1)	52 (26.4)	30 (13.3)	-7.2	(-13.5, -0.9)	0.0100
tBPAR	11 (4.8)	42 (19.9)	18 (7.7)	-2.9	(-7.9, 2.2)	0.2031
[redacted]						

P-value and 97.5% CI are obtained using Kaplan-Meier (KM) probability estimates of event rates and standard error derived based on Greenwood's formula.

Source: Table 14.2-2.1d

Table 12 shows that the difference in KM estimates between EVR+rTAC and TAC was 2.2% in favour of EVR+rTAC, CI (-8.8% to 4.4%). This means that the composite endpoint (tBPAR, graft loss or death) had fewer observations in the EVR+rTAC than in the TAC arm of the trial and that EVR+rTAC is statistically non-inferior to TAC in terms of the composite efficacy outcome at 24 months. Of note is that during the second year the relative increase in events in the TAC arm of the trial (6 events) was smaller than the relative increase in events reported in the EVR+rTAC arm (8 events).

Table 13 reports the statistical comparison between treatment arms for KM incidence rates of graft loss or death. Results from this analysis indicate that at 24 months EVR+rTAC is statistically non-inferior to TAC. However, similarly to the 12-month analysis, the estimated difference between treatment arms for the KM of graft loss or death favours the TAC arm, where less events occurred.

Table 14 presents a summary of the incidence rates for primary and secondary outcomes across treatment arms. Of note is that the p-values presented for the primary outcomes are not for the non-inferiority hypothesis, but for an equivalence hypothesis. Therefore what is tested is if the outcomes are statistically different (p-value<0.05) or not. Similarly to the 12-month analysis it can be observed that graft loss and death were not statistically different across the treatment arms, with p-values>0.05. Interestingly, by the end of the second year, treated acute rejection and tBPAR were also not statistically different across the EVR+rTAC and the TAC arm.

In Table 15 we present the summary outcomes reported for the 36 months.

Table 15. KM incidence rates of efficacy endpoints at Month 36 (ITT population)

Efficacy variable	EVR+Reduced TAC N=245 n (KM %)	TAC Elimination N=231 n (KM %)	TAC Control N=243 n (KM %)	Diff Risk (KM%)	97.5% CI	P-value
Composite efficacy failure: tBPAR, GL or D	26 (11.5)	56 (27.5)	32 (14.6)	-3.2	(-10.5, 4.2)	0.3337
[redacted]						
Graft loss (GL)	9 (3.9)	6 (2.8)	8 (4.0)	-0.0	(-4.3, 4.3)	0.9873
Death (D)	14 (6.5)	15 (7.3)	10 (4.4)	2.1	(-2.9, 7.0)	0.3483
[redacted]						
BPAR	15 (7.3)	53 (26.8)	34 (17.7)	-10.4	(-19.0, -1.9)	0.0064
tBPAR	11 (4.8)	43 (21.5)	20 (9.2)	-4.4	(-9.9, 1.2)	0.0762
[redacted]						

1. AR = Acute rejection; tAR = Treated acute rejection; BPAR = biopsy proven acute rejection; tBPAR = treated biopsy proven acute rejection.

2) P-values and 97.5% CI are obtained using Kaplan-Meier (KM) probability estimates of event rates and standard error derived based on Greenwood's formula.

Source: Table 14.2-1.2a

The 36-month CSR (Rauer et al, 2014) reported that the methods used for the statistical analysis undertaken for these data was similar to one reported in the Hexham et al, 2011 study (12 months data).

Table 15 indicates that fewer events were observed in the EVR+rTAC arm with respect to the composite efficacy endpoint. However, it is not clear if the p-value presented is for the non-inferiority test. Similarly to the 2-year analysis, the incidence of graft loss, death and tBPAR (considered separately) is not statistically different across treatment arms.

Figure 1 and Figure 2 present the KM curves for the composite efficacy measure and for the graft loss or death outcome, respectively, from randomization up to 36 months.

Figure 1. Kaplan-Meier plot of composite efficacy failure at Month 36 since randomization

[redacted]

Figure 2. KM for the proportion of patients free from graft loss or death (ITT population -36 months)

[redacted]

Looking at the KM plots in Figure 1 and Figure 2, it is interesting to observe how the TAC elimination arm results change so drastically when tBPAR is excluded from the analysis. When tBPAR is included as an outcome, together with graft loss or death (Figure 1) the KM curve of the TAC elimination shows that EVR monotherapy (the active drug on the elimination arm) performs a lot worse than EVR+rTAC or standard TAC in preventing tBPAR events. However, when tBPAR events are excluded from the analysis and only graft loss and death remain, EVR monotherapy seems to perform as good as the comparator drugs. This reinforces the point that the effectiveness of EVR monotherapy is largely dependent on the choice of clinical outcomes and whether these include acute rejection episodes or graft losses.

Subgroup baseline results

Figure 3 presents the subgroup analysis of incidence rates for the primary clinical endpoint (tBPAR, graft loss or death) undertaken in the Hexham et al, 2011 study. The subgroups considered were defined according to age, gender, race, eGFR, HCV status, MELD score and cause of ESLD.

Figure 3. Subgroup analysis at 12 months

[redacted]

The statistical analysis underkaten shows no statistically significant relationships between the subgroup factors and the composite efficacy outcome.

In the CSR it is stated that *the overall pattern of the composite event rates within subgroups (age, gender, race, eGFR, HCV status, MELD score, cause of ESLD) was similar to that observed in the general population.*

The two-year subgroup analysis led to the same conclusions as those inferred from the 12-month analysis.

Subgroup analysis was not underaken for the 36 months follow-up study (Rauer et al, 2014)

Efficacy results for renal outcomes

The ERG also used the CSR documents to report the main efficacy results for renal outcomes.

Table 16 and Table 17 present the comparison between treatment groups for the eGFR levels at 12, and 24 months respectively.

Table 16. eGFR (ITT population at 12 months)

			Difference vs. Control			
Treatment	**N**	**LS Mean (SE)**	**LSM Mean (SE)**	**97.5% CI**	**P value(1)**	**P value(2)**
EVR+Reduced TAC	244	-2.23 (1.54)	8.50 [redacted]	(3.74, 13.27)	<0.001	<0.001
TAC Elimination	231	-1.51 (1.58)	9.22 [redacted]	[redacted]		
TAC Control	243	-10.73 (1.54)				

1. Least squares means, 97.5% confidence intervals, and p-values are from an ANCOVA model containing treatment and HCV status as factors, and baseline eGFR as a covariate.
2. Imputation rule of missing Month 12 eGFR (MDRD4) values: 1) use the last available value before randomization for patients with no post-randomization eGFR; 2) use the minimal value if the last value is observed between randomization and Month 6; 3) use the minimal value between Month 6 and Month 12 if the last value is observed at or after Month 6. 4) use 15 mL/min/1.73m^2 if any patient was on dialysis after randomization.
3. Pvalue (1): Non-inferiority test with NI margin = -6 mL/min/1.73m^2, at one-sided 0.0125 level.
4. Pvalue (2): Superiority test at two-sided 0.025 levels.
Source: Table 14.3-2.1a

The difference between the EVR+rTAC arm and the TAC arm in mean GFR at month 12 was 8.50 mL/min/1.73m2 in favour of EVR+rTAC. Statistical non-inferiority and superiority of EVR+rTAC arm versus TAC was demonstrated [redacted]

Table 17. eGFR (ITT population at 24 months)

				Difference vs. Control		
Treatment	**N**	**LS Mean (SE)**	**LS Mean (SE)**	**97.5% CI**	**p-value(1)**	**p-value(2)**
EVR+Reduced TAC	245	[redacted]	6.66 [redacted]	(1.90, 11.42)	[redacted]	0.0018
TAC Elimination	231		10.41 [redacted]	(5.58, 15.25)		
TAC Control	243					

Least squares means, 97.5% confidence intervals, and p-values are from an ANCOVA model containing treatment and HCV status as factors, and baseline eGFR as a covariate.
Missing Month 24 eGFR (MDRD4) values are imputed (for details see Section 9.7.6.2 –renal function).
p-value(1): one-sided non-inferiority test with NI margin = -6 mL/min/1.73m^2.
p-value(2): Two-sided superiority test.
Source: Table 14.3-2.1a

The change in estimated GFR (MDRD4 formula) at month 24 showed a greater decrease from baseline to 24 months in the TAC control group than in the EVR+rTAC group, with an LS mean treatment difference of 6.66 mL/min/1.73m2. Statistical non-inferiority and superiority of EVR+rTAC arm versus TAC was demonstrated (p<0.0001; 97.5% CI: 1.90, 11.42).

Superior renal function was maintained at month 36 for the EVR arm with 16.7 mL/min/1.73m2 and 23.2 mL/min/1.73m2 differences in mean on-treatment eGFR values for EVR+rTAC and TAC elimination arms respectively vs. standard TAC.

Subgroup baseline results

The CSR reported that the calculated eGFR (MDRD4) for all the subgroups (according to age, gender, race, region, hepatitis C serology status, eGFR strata, MELD categories, ESLD) showed trends over time to be similar to the population as a whole at 6 and 12 months. Treatment comparisons for change of eGFR (MDRD4) from randomization visit showed that EVR+rTAC was statistically superior to TAC at all time-points ($p<0.0001$) by Wilcoxon rank-sum test.

Subgroup analysis for 24 and 36 months showed similar results to the 12-month analysis.

HCC outcomes

The CSR reported that recurrence of HCC during the first 12 months of study was observed for one patient in the TAC elimination arm and for one patient in the TAC control arm. The company reported that neither events were suspected to be study drug related.

Recurrence of HCC during the 24 months of study occurred in five EVR+rTAC patients, six TAC elimination patients and four patients in the TAC control arm.

A total of two patients experienced recurrence of HCC in the extension period of the study (from month 24 in the main trial to month 48 time). The company did not suspect a causal relationship between the event and the study medication.

HCV outcomes

EVR+rTAC achieved comparable results to standard TAC, with no negative impact on the evolution of viral replication but potentially positive impact on progression of HCV related fibrosis.

The CSR reported that the overall number of patients with HCV infection were few in the extension study with only 2.8% reporting HCV at 36 months.

- NMA

NMA results originally presented by the company

The NMA compared standard TAC, AZA+CIC, AZA+TAC, MMF+CIC, MMF+TAC, MMF+rTAC and EVR+rTAC. The NMA results were presented as a consistency model, where direct and indirect evidence are assumed consistent for any ‘closed loops’ in the evidence network. An inconsistency model was also presented where data were available. Inconsistency models only use the direct evidence available.

The NMA results for all 13 outcomes were presented by the company in Tables 23 to 39 in the submission. These were provided as odds ratios (ORs) alongside credible intervals.

The data ultimately used in the economic model, didn't consist of OR but instead on probability of events. The ERG requested that the company explained how these were derived, to which the company replied by presenting what was considered *the absolute results for the NMA outcomes.* These were basically the final values used in the model, for which the ERG requested initial clarification. The ERG are aware that these absolute estimates can be obtained through a WinBUGS command. Nonetheless, in order to be able to compute these in the software package, baseline effectiveness data need to be inputted for the comparator treatments. Novartis do not provide these data in their submission, nor make reference to it. Again, because the ERG didn't get access to the WinBUGS code using the actual submission data, we could not verify this.

The absolute NMA results are now presented.

Absolute estimate results for NMA

The ERG present results for the following estimates:

- Overall survival at 12 and 24 months
- Graft survival at 12 and 24 months
- tBPAR free at 3, 6 and 12 months
- Expected absolute results for renal function at 12 months

However, as only the probability of being tBPAR free and renal function were used in the economic model, we mainly focus on these outcomes.

The company also presented the expected absolute results for withdrawals due to adverse events at 12 months (hypertension and diabetes) and the expected absolute results for infections and HCV recurrent 12 months. The reader should consult the clarification document for details on these.

For overall survival and graft loss, Novartis have used different studies to report the 12 month analysis and the 24 month analysis. This was also the case for time BPAR free, where studies used in the network varied for 3,6 and 12 months. Due to incompatibilities, only studies reporting a 12 and 24 month analysis for overall survival and graft loss were used. The same was applied to time BPAR free, where only the studies reporting a 3, 6 and 12 month analysis were used with goal to increase compatibility.

The ERG interpreted that by "compatibility" the company meant the decreasing trend observed in results across the various time frames. Novartis reported that the expected absolute point estimates for overall survival at 12 and at 24 months were compatible for EVR + reduced TAC (87.1% [95%CrI: 71.1; 94.9] at 12 months and 85.3% [95%CrI: 72.5; 92.9] at 24 months), but that incompatibility issues were observed for for MMF + CIC (78.0% [95%CrI: 38.3; 93.6] at 12 months and 88.9% [95%CrI: 43.6; 98.8] at 24 months*).*

To increase compatibility, the company decided to substantially reduce the amount of studies included in the estimation of different outcomes by dropping studies initially included in the NMA. Consequently, direct evidence was lost for specific time points thus reducing the overall sample size of the evidence base.

Figure 4 presents the overall survival at 12 and 24 months along with the corresponding credible intervals and ranking.

The wide intervals for MMF+CIC and AZA+CIC at 12 months reflect the high level of uncertainty for these results. The ERG note that EVR+rTAC was ranked as 5th and 4th amongst the interventions at 12 and 24 months respectively.

Figure 4. Expected absolute results for overall survival at 12 and 24 months (fixed effects model)

Intervention	Absolute estimate 95% CrI Ranking
standard TAC	91.30% 88.50% 93.70% 3
AZA + CIC	76.10% 36.60% 92.60% 7
AZA + standard TAC	92.80% 88.40% 95.80% 2
MMF + CIC	78.00% 38.30% 93.60% 6
MMF + standard TAC	92.00% 88.10% 94.90% 2
MMF + reduced TAC	91.10% 83.60% 95.30% 3
EVR + reduced TAC	87.10% 71.10% 94.90% 5

Intervention	absolute estimate 95% CrI Ranking
standard TAC	87.40% 84.10% 90.20% 3
AZA + standard TAC	85.80% 76.70% 91.80% 4
MMF + CIC	88.90% 43.60% 98.80% 2
MMF + standard TAC	88.40% 83.80% 92.00% 2
EVR + reduced TAC	85.30% 72.50% 92.90% 4

Figure 5 presents the absolute results for graft survival at 12 and 24 months, along with corresponding 95% credible intervals and ranking of therapies.

The wide intervals observed for MMF+CIC at 12 and 24 months and for AZA+CIC at 12 months reflect the high levels of uncertainty in these results.

The ERG also note that EVR+rTAC effectiveness in terms of graft survival was ranked in the bottom when compared with other therapies at 12 months and it ranked number 4 for graft survival at 24 months.

Figure 5. Expected absolute results for graft survival at 12 and 24 months (fixed effects models)

Intervention	absolute estimate 95% CrI Ranking
standard TAC	87.90% 84.80% 90.70% 3
AZA + CIC	70.20% 37.50% 88.20% 6
AZA + standard TAC	89.50% 84.20% 93.20% 2
MMF + CIC	71.10% 37.70% 89.00% 6
MMF + standard TAC	90.50% 86.30% 93.70% 2
MMF + reduced TAC	89.20% 79.60% 94.60% 3
EVR + reduced TAC	80.40% 63.40% 90.60% 5

Intervention	absolute estimate 95% CrI Ranking
standard TAC	82.60% 79.10% 85.80% 3
AZA + standard TAC	80.80% 70.60% 88.20% 4
MMF + CIC	86.00% 49.40% 97.50% 2
MMF + standard TAC	85.30% 80.00% 89.60% 2
EVR + reduced TAC	79.70% 67.30% 88.40% 4

Source: Novartis clarification document

Figure 6 presents absolute results for BPAR at 3, 6 and 12 months for the fixed effects models. In the original NMA analysis, 11 studies reported this outcome at 3 months, 7 studies reported the outcome at 6 months and a total of 9 studies reported outcomes at 12 months.

However, for the purposes of increasing compatibility between outcomes, Novartis only used studies which consistently included 3, 6 and 12 month analysis for 'time BPAR free'. Novartis report including only 2 studies (Boudjema 2011 and H2304) in the final analysis, which presented outcomes for all time periods (i.e. 3, 6 and 12 months). The ERG note that a larger network of studies would be required to incorporate all the comparators which are part of the BPAR 3, 6 and 12 month results. It is unknown which studies were used or how AZA was estimated in the expected results as this was not reported.

Novartis report that these 2 studies were then incorporated into calculating a weighted mean standard TAC value to establish absolute estimates. The ERG are not clear what is meant by this.

The company reported that even after reducing the number of studies included in the analysis the results for AZA+CIC still showed incompatibility. Furthermore, the wide intervals for AZA+CIC at 3, 6 and 12 months reflect the high uncertainty of these results. EVR+rTAC was ranked as the best therapeutic option amongst all therapies for time BPAR free.

Figure 6. BPAR free at 3, 6 and 12 months (fixed effect models)

3 months Intervention	absolute estimate 95% CrI Ranking		
standard TAC	76.7%	81.6%	86.0%
		6	
AZA + CIC	62.9%	79.4%	89.7%
		7	
AZA + standard TAC	80.2%	86.2%	90.6%
		4	
MMF + CIC	68.0%	84.0%	92.8%
		5	
MMF + standard TAC	82.7%	88.7%	92.9%
		3	
MMF + reduced TAC	85.4%	90.8%	94.5%
		2	
EVR + reduced TAC	93.5%	98.1%	99.7%
		1	

6 months Intervention	absolute estimate 95% CrI Ranking		
standard TAC	75.8%	80.4%	84.6%
		4	
AZA + CIC	27.9%	71.1%	93.4%
		7	
AZA + standard TAC	33.6%	78.7%	96.0%
		5	
MMF + CIC	37.3%	78.7%	95.4%
		5	
MMF + standard TAC	56.3%	84.6%	95.8%
		3	
MMF + reduced TAC	77.9%	83.8%	88.5%
		3	
EVR + reduced TAC	88.5%	94.4%	97.9%
		1	

12 months Intervention	absolute estimate 95% CrI Ranking		
standard TAC	72.0%	76.8%	81.2%
		4	
AZA + CIC	55.4%	72.3%	84.6%
		6	
AZA + standard TAC	65.3%	75.6%	83.7%
		5	
MMF + standard TAC	75.8%	83.4%	88.9%
		2	
MMF + reduced TAC	74.3%	80.6%	85.9%
		3	
EVR + reduced TAC	82.3%	89.5%	94.4%
		1	

Source: Novartis clarification documen

Table 18 presents absolute results for the renal function at 12 months, alongside the 95% credible intervals and ranking of therapies.

Table 18. Expected absolute results for renal function at 12 months (fixed effects model)

Intervention	***absolute estimate*** ***95% CrI*** ***Ranking***
standard TAC	-31.6 -32.3 -30.9 4
AZA + CIC	-14.5 -24.2 -4.9 1
MMF + reduced TAC	-28.2 -32.3 -24.1 3
EVR + reduced TAC	-23.1 -27.4 -18.7 2

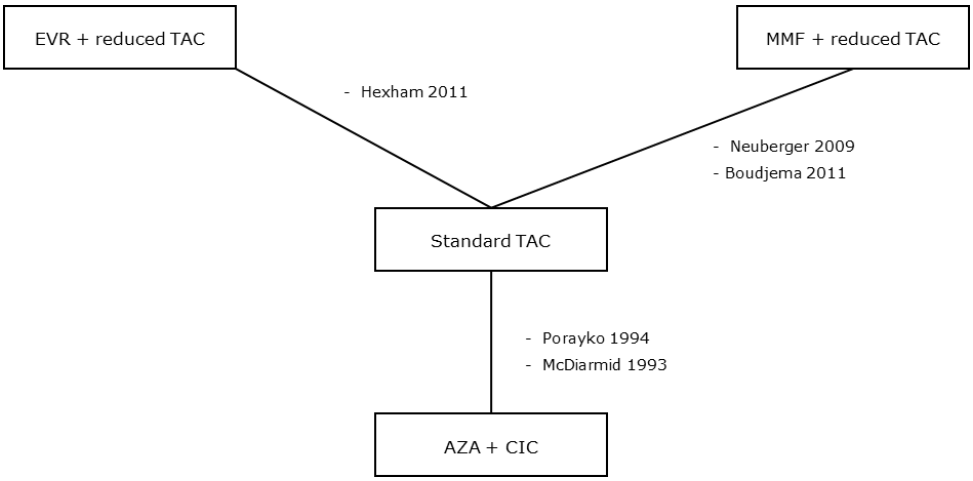
Source: Novartis clarification document

The renal results in Table 18 estimate that the therapy leading to the lowest decline in renal function at 12 months (i.e with the highest renal sparing effect) is AZA+CIC, followed by EVR+rTAC, MMF+rTAC and finally, standard TAC. However, the credible interval for AZA+CIC is considerably wide, indicating a high level of uncertainty.

In the original NMA analysis, Novartis used 5 studies, namely Hexham et al 2011, Neuberger et al 2009, Boudjema et al 2011, Porayko et al 1994 and McDiarmid et al 1993 at 12 months to estimate renal function at 12 months. In Figure 7 we present the network diagram for the outcome.

The studies reported the results as either eGFR or eCCR. In order to combine the results of the eGFR and eCCR measures in one analysis, effect size (ES) analysis was conducted. The difference between the least square mean at follow-up or the change from baseline for the baseline comparator versus the active treatment was divided by the pooled standard deviation (SD) so that both measurements were standardised and combined in one analysis. The ERG consider this method satisfactory and in line with NICE DSU 2.

Figure 7. Network for renal function at 12 months



Source: Novartis submission Appendix 14

Scenario analysis

Given that the company identified potential treatment modifiers (described above), the ERG requested clarification from Novartis as to why the relevant covariates were not taken into account with treatment-by-covariate interactions in a meta-regression model to try to reduce the impact of bias.

Novartis clarified that an attempt was made to use a meta-regression model, but due to the paucity of studies, covariate analyses could not be robustly performed. Therefore treatment effect modifiers were analysed through outcome-specific scenario analyses.

The NMA scenario analysis undertaken by the company consisted in removing specific trials from the network to assess the impact on the NMA results. The different scenarios considered were:

1. Grouping CNIs. In this scenario the treatment arms would be:

    - EVR + reduced TAC
    - AZA + standard CNI
    - MMF + standard CNI
    - MMF + reduced TAC
    - Standard TAC.

2. Excluding studies that included treatment arms with steroids < 3 months

3. Excluding studies where TAC could be reclassified. The company acknowledged that most of the studies included in the NMA and that were considered standard TAC, would more accurately fall under different categories (like for example reduced TAC) and vice-versa. Therefore they have eliminate these studies from the analysis.
4. Excluding studies with HCC imbalances at baseline
5. Excluding studies in which patients with diabetes are excluded at baseline
6. Excluding studies that use a comparator as a rescue medication
7. Including only eGFR measures for renal function CFB at 12 months

Results of scenario analyses

The results presented by the company are extremely disappointing. No actual results were provided (neither in the submission or other documents) and Novartis simply reported that the only scenario which had an impact on BPAR free at 3 and 6 months was scenario 5 and that all other outcomes were insensitive to the scenario analysis.

The scenario analysis undertaken lacks transparency and is not informative.

- ## Adverse Events

Although a literature search was carried out, no relevant studies were identified that included relevant AEs. Table 19, shows the key adverse events reported across treatment groups in H2304.

Based on the results from Table 19, the relative risk of diarrhoea was higher in the TAC control arm than the EVR+rTAC, however, for headaches, lipid changes and infections, the relative risk is higher in the EVR+TAC group compared to the TAC control. At 24 months, this becomes increasingly worse or is not reported. The company provided no statistical analysis for the differences shown.

Table 19. Key adverse events across randomised groups from H2304

System organ / class / adverse events	12 months			24 months		
	EVR+ rTAC (N = 245) n (%)	***TAC control (N = 241) n (%)***	***Relative risk, % (95% CI)***	***EVR+ rTAC (N = 245)***	***TAC control (N = 241)***	***Relative risk, % (95% CI)***
Diarrhoea	*47 (19.2)*	*50 (20.7)*	*0.92 (0.65-1.32)*	[redacted]	[redacted]	[redacted]
Headache	*47 (19.2)*	*46 (19.1)*	*1.01% (0.70-1.45)*	[redacted]	[redacted]	[redacted]
Hypertension	*42 (17.1)*	*38 (15.8)*	*NR*	[redacted]	[redacted]	[redacted]

Wound healing/ biliary leaks	*27 (11)*	*19 (7.9)*	*1.4 (0.80-2.45)*	*27 (11)*	*20 (8.3)*	*2.8 (2.5-8.0)*
New onset diabetes mellitus	*48 (19.6)*	*40 (16.6)*	*NR*	*51 (20.8)*	*40 (16.5)*	*4.3 (2.6-11.2)*
Lipid changes	*58 (23.7)*	*23 (9.5)*	*2.48 (1.58-3.89)*	*66 (26.9)*	*28 (11.6)*	*15.4 (8.5- 22.2)*
Infections	*123 (50.2)*	*105 (43.6)*	*1.15 (0.95-1.39)*	[redacted]	[redacted]	[redacted]
Renal failure	*15 (6.1)*	*21 (8.7)*	*NR*	[redacted]	[redacted]	[redacted]
CI, confidence interval; NR, not reported *EVR + rTAC, everolimus with reduced dose tacrolimus and corticosteroids* *TAC control, tacrolimus with corticosteroids control arm* *Adapted from European Public Assessment Reports published by the European Medicines Agency*						

Source: Novartis submission Table 43

4.2.2 Summary of clinical effectiveness

It is likely that the systematic review on the literature for clinical effectiveness undertaken by Novartis contains all the relevant studies. The evidence submitted generally reflect the decision problem outlined in the final scope of the submission.

EVR+rTAC effectiveness relied on evidence drawn from the H2304 trial, which quality is good. It seems that the composite primary endpoint of the trial (tBPAR, graft loss and death) combines the effectiveness of two outcomes which work in different directions. Whilst graft loss or death seem to favour the TAC arm of the trial, the addition of tBPAR to the composite endpoint favours the EVR+rTAC arm. This casts doubt on the appropriateness of the clinical endpoints as previously discussed.

The trough levels of reduced TAC in the trial are above of what would be considered as reduced levels in the UK clinical practice.

The ERG found a significant limitation in the NMA presented by Novartis as the data included in the WinBUGS codes do not relate to the submission data in any way. This poses a major limitation as the ERG could not verify which data were used for the analysis of specific outcomes. This represents a concern as the ERG are not clear which studies have been included in the NMA analysis for the tBPAR outcome, due to lack of clarity and transparency in the submission.

The scenario analysis undertaken for the NMA trials lacks transparency and is not informative.

5.0 Economic evaluation

In this chapter, we assess the cost-effectiveness analysis submitted by Novartis.

We start with a summary of the systemic review of cost-effectiveness studies presented by Novartis and the methods used in the economic evaluation (Section 5.1). Then we present a critique of the methods they used (Section 5.2). This is followed by a description of Novartis' results (Section 5.3) and our comment on their validity (Section 5.4).

5.1 Overview of company's economic evaluation

5.1.1 Summary of Novartis' systematic review of cost-effectiveness studies

- #### Description of company's search strategy and comment on whether the search strategy was appropriate

Following satisfactory answers to question raised in clarification, the ERG are content to accept the cost-effectiveness searches ran by Novartis.

The search strategy takes the following form and it was last updated in August 2014, so the results are considered current for this submission:

(Terms for liver transplant or hepatic transplant or graft) AND (terms for Everolimus) AND (a standard economics/costs search filter)

The company searched all of the required resources for this section of the submission.

The ERG point to the following limitations identified on the searches undertaken:

The company, following clinical advice, used a date limit on their cost-effectiveness searches of 1990-Current.

The ERG asked the company to confirm how the MEDLINE c/e searches were combined on page 271 of the manufacture's submission. The company confirmed that there had been a copy and paste error in their submission. The searches should read:

Line 40: '4 AND 7 AND 39'
Line 41 date limit (1990-current)

- #### Search results

A range of studies were identified and their relevance assessed according to the inclusion/exclusion criteria described in Table 44 of Novartis' submission.

While at the primary review the majority of the studies failed to meet the inclusion criteria, at the secondary review, no studies met the inclusion criteria. Therefore, no relevant cost-effectiveness studies were found. For this reason, a de-novo analysis was undertaken.

5.1.2 Novartis' economic model submitted to NICE

We now turn to the economic evaluation that Novartis presented to NICE. Novartis report costs per QALY estimates for EVR+rTAC with concomitant corticosteroids for the initial 6 months of therapy compared to MMF+TAC and AZA+TAC both with concomitant corticosteroids for the initial 6 months of therapy.

Superseded
See Erratum

The model was built in Microsoft Excel©. Here, we summarise the main features of the model.

Throughout their submission Novartis acknowledge that *the advantage of competing immunosuppression regimens is not through limiting rejection (as this is managed very well currently) but it is obtained by improvements in side-effects such as impaired renal functioning*. Therefore, the economic analysis undertaken by the company intends to demonstrate the benefit of EVR in terms of its kidney sparing effect, whilst guaranteeing at least the same effectiveness as the comparator drugs in terms of graft survival.

- #### Model structure

Novartis' cost-effectiveness model was developed as a patient simulation model. The structure of the economic model[3], illustrated in Figure 8, includes a core hepatic model and a renal sub-model and is reported to be appropriate and reflective of the clinical pathway of immunosuppression therapy after liver transplantation.

Novartis state that the use of a patient simulation model is in line with the DSU technical support document 15 (Davis et al, 2014) recommendations as the technical report suggests this modelling approach is appropriate when the patient flow is determined by time since last event or history of previous event.

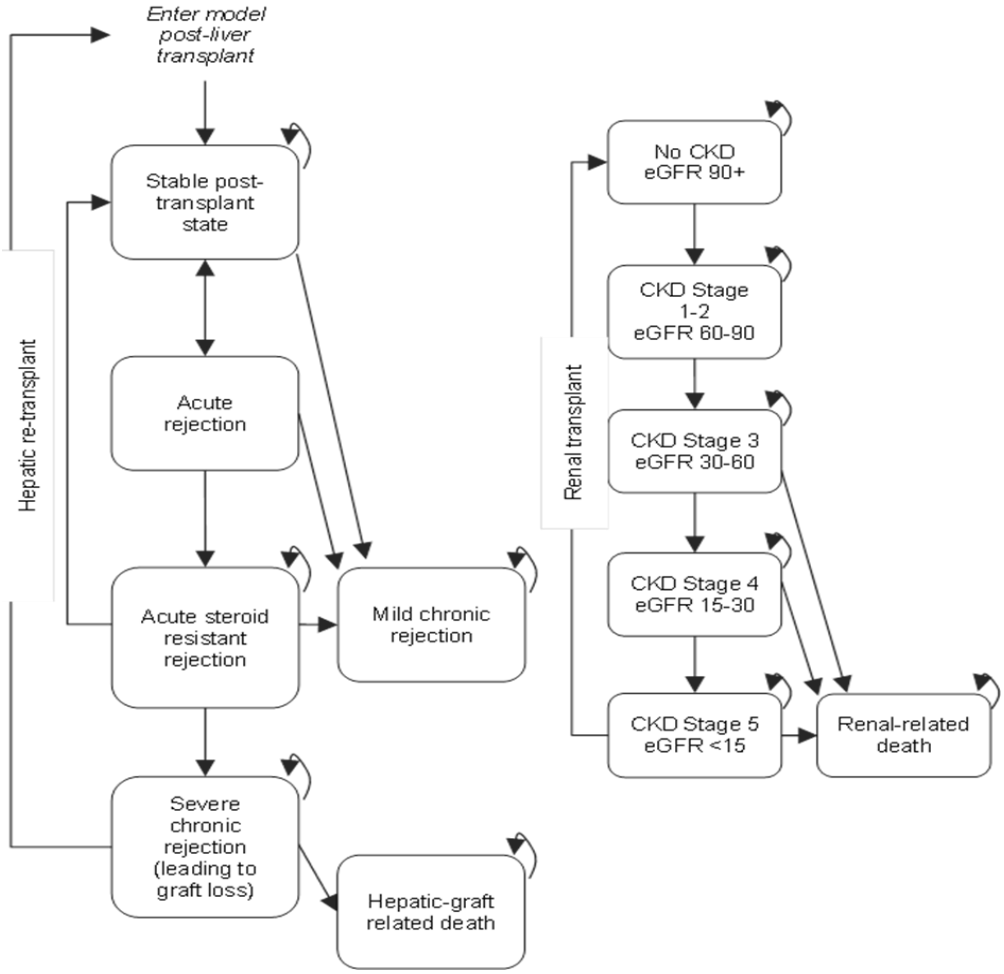
Novartis also explain that the decision to capture the renal sparing effect associated with EVR+rTAC (demonstrated in the H2304 trial) through a renal sub-model *was considered important because the treatment effect* [survival of liver graft] *has an impact on more than one aspect of a patients' health.* Additionally it is mentioned that *ISPOR good research practice guidelines encourage the design option of using sub-models to simplify the model structure (Karnon et al., 2012).*

Furthermore, the company reports that a patient level simulation approach facilitates *between-model calculations* involving the hepatic core model and the renal sub-model. However it adds that *in the patient simulation, events that occur in the core* [hepatic] *model do not impact on*

[3] The "economic model" refers to the core hepatic model and the renal sub-model throughout the rest of the document.

progression through the renal sub-model. The exception to this is death, which can occur as a result of progression in either model, and is considered an absorbing state in both models.

Figure 8. Core hepatic cost-effectiveness model and renal sub-model



Source: Novartis Submission Figure 10

Novartis provide a brief description of the 6 health states included in the hepatic core model. These are presented in Table 20.

Table 20. Health states considered in the hepatic core model

Health state	Description
Stable post-transplant state (SPT state)	Patients enter the model four weeks after their transplant in this state which reflects being stable on immunosuppressive regimens with no clinical signs of rejection. Patients who successfully treat acute rejection episodes return to this state.

Acute rejection (AR)	Patients with acute rejection episodes (as diagnosed by biopsy) enter this state and can either return to the stable post-transplant state through successful high-dose steroid treatment or progresses to the steroid-resistant acute rejection state after 3 months.
Acute steroid resistant rejection (ASRR)	If steroid treatment is unsuccessful after 3 months, patients enter this state where they undergo second line anti-lymphocyte therapy.
Severe chronic rejection (leading to graft loss) (SCR)	If a patient fails both first line (high-dose steroid) and second line (anti-lymphocyte) therapy for acute rejection, a patient enters a state of severe chronic rejection. Severe chronic rejection will lead to failure of the transplanted organ and the patient will require a further transplant.
Mild chronic rejection (MCR)	After year 1 post-transplant, a patient can develop a mild form of chronic rejection from any state except severe chronic rejection and death. This state is largely asymptomatic. It is not clear what causes this state of rejection, and a patient can also recover naturally and return to a stable post-transplant state with no rejection (Hirschfield et al., 2009).
Death	Patients have increased mortality risk in a state of severe chronic rejection (leading to graft loss) and this is captured by the liver-related death state in the model. Natural (background) mortality risk is applied at all stages of the model.

Source: Adapted from Novartis Submission Table 45.

In the renal sub-model, renal dysfunction was determined through the decrease in patients' baseline level of eGFR. Depending on the eGFR level the patient can develop chronic kidney disease (CKD) as per the thresholds reported in Table 21.

On page 142 of their submission, Novartis describe the 6 health states included in the renal sub-model. These are reported to have been taken from the NICE guidance on CKD progression and are presented in Table 21.

Table 21. Health states in renal sub-model

Health state	***eGFR***	***Definition (from the NICE guideline for CKD)***
No CKD	*≥90 mL/min/1.73m2*	*Normal or increased GFR, with other evidence of kidney damage.*
Stages 1-2	*60-89 mL/min/1.73m2*	*Slight decrease in GFR, with other evidence of kidney damage.*
Stage 3	*30-59 mL/min/1.73m2*	*Moderate decrease in GFR, with or without other evidence of kidney damage.*
Stage 4	*15-29 mL/min/1.73m2*	*Severe decrease in GFR, with or without other evidence of kidney damage.*
Stage 5	*0-15 mL/min/1.73m2*	*Established renal failure, in need of dialysis or transplant.*
Death		*Patients have increased mortality risk in CKD 5 and this is captured in the renal-related death state in the model. Natural (background) mortality risk*

		is applied at all stages of the model.

Source: Novartis Submission Table 46.

Novartis state that the health states included in the renal sub-model accurately reflect the relevant health services costs and effects arising from post-transplant complications. It is reported that *once eGFR falls below 60 mL/min/1.73 m^2 complications should be treated* and that *approximately 98% of patients with kidney failure begin dialysis when their eGFR is less than 15 mL/min/1.73 m^2*.

Even though the health state is not presented in Figure 8, a natural mortality state was also considered. The company reports that a background mortality rate was applied to patients in all states of the model.

When comparing their model structure with the post-liver transplantation clinical pathway, Novartis note that if post-transplant immunosuppression therapy fails and graft rejection occurs, then it can be categorized into hyperacute, acute (cellular) or chronic (ductopenic) rejection. Novartis decided not to include the hyperacute rejection state in their economic model as it is reported to occur within the first minutes or hours post-transplantation. Therefore Novartis conclude that EVR+reduced TAC (administered to patients 30 days after transplantation) would not have an impact on hyperacute rejection, making the health state irrelevant to the decision problem.

The cycle length in the economic model is 3 months and a half-cycle correction was not applied. The company justify the choice of cycle length with clinical expert opinion which stated *that most acute rejection occurs after 3 months post transplantation.*

The time horizon considered in the economic model was lifetime (the model was ran for 320 cycles - 80 years).

- ### Treatment effectiveness within submission

Treatment effectiveness within the model (both the core hepatic model and the renal sub-model) works essentially through transition probabilities between some of the health states presented in the previous section.

In this subsection we briefly describe how transition probabilities between health states were estimated within the different arms of the main economic model. We provide further detail in Section 5.2.

Core hepatic model

In the hepatic model, treatment effectiveness only impacted the patients' transition from the stable post-transplant (SPT) to the acute rejection (AR) state. This means that in the economic analysis, the type of immunosuppressive treatment the patient receives (i.e. EVR+rTAC or any

of the comparator regimens) is only relevant to determine the likelihood of the patient experiencing an AR episode.

All other transition probabilities were assumed constant and independent of the immunosuppressive treatment that patients receive.

In their submission the company reports that the probability of progressing from the SPT to the AR state was determined by the probability of BPAR at 3, 6 and 12 months for each comparator, calculated in the NMA. However, no further details were given as to how the NMA BPAR data (originally presented as odd ratios) were used to derive time dependant transition probabilities.

Upon the ERG's request for further detail on the aforementioned calculations, Novartis provided the tables shown in Figure 9 below (also presented in Section 4 of this report) and reported that these are the absolute results for the outcomes presented in the NMA (corresponding to Tables 23-37 in the submission). The values presented are the probabilities of being tBPAR free at different points in time.

The ERG present the remaining transition probabilities across health states in the hepatic model in Table 22. The ERG report the values used in the economic model for the 3-month cycles. The company reported some of the parameters as annual estimates (instead of the cycle parameters used in the model). We explore this further in Section 5.2.

Figure 9. NMA tBPAR estimates used for transition probabilities in the hepatic model

3 months Intervention	absolute estimate 95% CrI Ranking		
standard TAC	76.7%	81.6%	86.0%
		6	
AZA + CIC	62.9%	79.4%	89.7%
		7	
AZA + standard TAC	80.2%	86.2%	90.6%
		4	
MMF + CIC	68.0%	84.0%	92.8%
		5	
MMF + standard TAC	82.7%	88.7%	92.9%
		3	
MMF + reduced TAC	85.4%	90.8%	94.5%
		2	
EVR + reduced TAC	93.5%	98.1%	99.7%
		1	

6 months Intervention	absolute estimate 95% CrI Ranking		
standard TAC	75.8%	80.4%	84.6%
		4	
AZA + CIC	27.9%	71.1%	93.4%
		7	
AZA + standard TAC	33.6%	78.7%	96.0%
		5	
MMF + CIC	37.3%	78.7%	95.4%
		5	
MMF + standard TAC	56.3%	84.6%	95.8%
		3	
MMF + reduced TAC	77.9%	83.8%	88.5%
		3	
EVR + reduced TAC	88.5%	94.4%	97.9%
		1	

12 months Intervention	absolute estimate 95% CrI Ranking		
standard TAC	72.0%	76.8%	81.2%
		4	
AZA + CIC	55.4%	72.3%	84.6%
		6	
AZA + standard TAC	65.3%	75.6%	83.7%
		5	
MMF + standard TAC	75.8%	83.4%	88.9%
		2	
MMF + reduced TAC	74.3%	80.6%	85.9%
		3	
EVR + reduced TAC	82.3%	89.5%	94.4%
		1	

Table 22. Three-month transition probabilities used in the base case hepatic model across treatment arms

From health state	To health state		Probabilities (all regimens)	Assumption	Source
Stable post-transplant state	Stable post-transplant state	Time and treatment dependant	Dependant on p (acute rejection)	Calculated as a residual value	1 - P(leaving state)
	Acute rejection	Time and treatment dependant	See Figure 9.	Calculated from tBPAR data in the NMA	NMA
	Acute steroid resistant rejection	Not time dependant	0%	Patients in the stable state can only move to the AR state	-
	Mild chronic rejection	Before 1 year	0%	The company assumed that this transition can only occur 1 year after transplant. The company reported this value to be 4% However the final probability used in the model is 1%. This is further explored in Section 5.2.	Clinical opinion
		After 1 year	1.0%		
	Death	Depends on the patient age	Life table	Natural mortality only	ONS, 2012
Acute rejection state	Stable post-transplant state	Not time dependant	98.1%	The probability of a 3-day steroid treatment resolving an AR episode (Aydogan et al, 2010) is 86.2%. However the final probability used in the model is 98.1%. This is further explored in Section 5.2.	Aydogan et al, 2010
	Acute rejection state		0%	Patients are not allowed to stay in the cycle for longer than 3 months (1 model cycle).	-
	Acute steroid resistant rejection		1.9%	Patients who fail the steroid treatment move on to this state	1 - P(leaving state)

	Mild chronic rejection		0%	Only patients in the SPT are allowed to transit to the MCR state	-
	Graft loss (severe chronic rejection)		0%	-	-
	Death	Depends on the patient age	Life table	Natural mortality only	ONS, 2012
Acute steroid resistant rejection state	Stable post-transplant state	Not time dependant	83.3%	Effectiveness of 1 course of anti-lymphocyte therapy	Aydogan et al, 2010
	Acute rejection state		0%	-	-
	Acute steroid resistant rejection		0%	Patients are not allowed to stay in the cycle for longer than 3 months (1 model cycle).	-
	Mild chronic rejection		0%	Only patients in the SPT are allowed to transit to the MCR state	-
	Graft loss (severe chronic rejection)		16.7%	Progression to chronic severe rejection if steroid therapy failed	Aydogan et al, 2010
	Liver-related death	Depends on the patient age	Life table	Natural mortality only	ONS, 2012
Mild chronic rejection state	Stable post-transplant state	Not time dependant	0%	Patients stay in this state unless they die	Aydogan et al, 2010
	Acute rejection state		0%		
	Acute steroid resistant rejection		0%		
	Mild chronic rejection		100%		
	Graft loss (severe chronic rejection)		0%		

	Liver-related death	Depends on the patient age	Life table	Natural mortality only	ONS, 2012
Severe chronic rejection (graft loss)	Stable post-transplant (via re-transplant)	First 3 months (i.e 1 cycle) in the state:	0%	Patients are assumed to stay in this state for 2 cycles (6 months) and are re-transplanted after this period.	Organ Donation, 2014
		After 6 months (i.e. 2 cycles) in the state:	99.3%		
	Acute rejection state	Not time dependant	0%	All patients in this state are re-transplanted after 6 months and go back to the SPT (re-transplant)	Organ Donation, 2014
	Acute steroid resistant rejection				
	Mild chronic rejection				
	Graft loss (severe chronic rejection)	First 3 months (i.e 1 cycle) in the state:	97.0%	Patients are assumed to stay in this state for 2 cycles (6 months) and are re-transplanted after this period.	Organ Donation, 2014
		After 6 months (i.e. 2 cycles) in the state:	0%		
	Death*	First 3 months in the state:	3.0%	12% of patients are assumed to die on the waiting list (annual estimate). However the final probability used in the model is 3%. This is further explored in Section 5.2.	Organ Donation, 2014
		After 3 months in the state:	0.7%	Probability of operative death at liver re-transplantation	Aylin et al, 2013

*Also includes background mortality rate as in other states

Renal sub-model

In the renal sub-model, treatment effectiveness has an impact on patient's renal functioning for the first year of the model. Renal dysfunction was measured through the decrease in patients' baseline level of eGFR.

The company explains that the renal complications resulting from immunosuppressive therapy are attributable only to CNIs (in this case TAC) and that the advantage of EVR in terms of its renal sparing effect comes from the reduced concomitant TAC dose use enabled by the drug. The nephrotoxicity of TAC during the first year of the economic model was captured through the

NMA estimates of the mean change in eGFR from baseline to 12 months, for the different treatment regimens.

The transition probabilities in the renal sub-model define the progression of CKD for each patient. The patient stating level of CKD is randomly generated through a simulation thus it is not possible to report the transition probabilities. However, the decrease in renal function (dependant on the treatment regimen) applied to the baseline CKD stage is fixed and is reported in Table 23. So for example, if a patient is in the EVR+rTAC arm, they will see a reduction of 23.1 mL/min/1.73 m^2 in their eGFR levels, which will correspond to a specific stage of CKD, depending on the patient starting level.

Superseded
See Erratum

Novartis didn't provide details as to how the NMA eGFR data (originally presented as the difference in eGFR change from baseline) were used to derive the eGFR decrease at 12 months for the different treatment regimens used in the economic model, presented in Table 23.

Upon the ERG's request for further detail on how the company used the NMA data to obtain the estimates in Table 18, Novartis provided the tables shown in Table 24 and reported that these are the absolute results for the outcomes presented in the NMA (corresponding to tables 23-37 in the submission).

The company added that the values reported in Table 24 show that AZA+CIC is the more effective treatment in terms of preserving renal function at 12 months, followed by EVR+rTAC, MMF+rTAC and finally, standard TAC. It was also stated that the credible intervals show some overlap between therapies.

Comparing Table 24 with Table 23 it can be noted that there is a discrepancy between the NMA estimates and the estimates used in the economic model. More specifically, the estimates for the decrease in eGFR at 12 months used in the comparator arms of the economic model (Table 23) are based on the one reported for standard TAC in the NMA (Table 24).Novartis justify this by arguing that as ciclosporin has limited use in the UK (approximately 5% of current market - Novartis data on file 2014), and the comparators of interest are used in combination with standard dose tacrolimus, only the absolute value for standard dose tacrolimus was used in the economic model.

Table 23. Absolute eGFR decrease at 12 months as used in the renal sub-model

Variable	*Value*	*CI (distribution)*	*Reference to section in submission*
Everolimus + reduced dose tacrolimus			
eGFR decrease at 12 months	*-23.1 mL/min/1.73m^2*	*-27.4 to -18.7 (Log-normal distribution)*	*Absolute values derived from NMA results section 6.7*
Azathioprine + standard dose tacrolimus			
eGFR decrease at 12 months	*-31.6 mL/min/1.73m^2*	*-32.3 to -30.9 (Log-normal distribution)*	*NMA results section 6.7*
MMF + standard dose tacrolimus			
eGFR decrease at 12 months	*-31.6 mL/min/1.73m^2*	*-32.3 to -30.9 (Log-normal distribution)*	*NMA results section 6.7*
CI, confidence interval;			

Table 24. Expected absolute results for renal function at 12 months (NMA fixed effects model)

Intervention	*absolute estimate* *95% credible interval (CrI)* *Ranking*
standard TAC	***-31.6**** *-32.3* *-30.9* *4*
AZA + CIC	*-14.5* *-24.2* *-4.9* *1*
MMF + reduced TAC	*-28.2* *-32.3* *-24.1* *3*
EVR + reduced TAC	***-23.1**** *-27.4* *-18.7* *2*

*These are the values used in the economic analysis

Novartis state that the results from the NMA show that *everolimus plus reduced dose tacrolimus is expected to be more effective at reducing renal function decline against standard dose tacrolimus, and is expected to be comparable to azathioprine plus ciclosporin and MMF plus reduced dose tacrolimus*.

Novartis assumed that by the end of year 1, patients' renal function follows a natural progressive decline. This means that by the end of the first year, all transition probabilities become constant and renal function decline is independent of the immunosuppressive treatment received. These probabilities are presented in Table 25.

Table 25. Transition probabilities in the renal sub-model independent of treatment effectiveness

From health state	*To health state*	*Probability (all regimens)*	*Assumption*	*Reference / source*
No CKD	*No CKD*	*98.25%*	*1 - P(leaving state)*	*Calculation*
	CKD stages 1 to 2	*1.75%*	*Annual probability of progression converted to rates and adjusted for cycle length, then re-converted to probability*	*(La Mattina et al., 2011)*
	Death (non-kidney related)	*0.00%*	*Natural mortality is assumed and is already accounted for in the main hepatic model*	*(ONS, 2012)*
CKD stages 1 to 2	*CKD stages 1 to 2*	*90.76%*	*1 - P(leaving state)*	*Calculation*
	CKD stage 3	*8.50%*	*Annual probability of progression converted to rates and adjusted for cycle length, then re-converted to probability*	*(La Mattina et al., 2011)*
	Kidney-related death	*0.74%*	*Added mortality risk assumed in CKD stage 2*	*(Keith et al., 2004)*
CKD stage 3	*CKD stage 3*	*84.79%*	*1 - P(leaving state)*	*Calculation*
	CKD stage 4	*14.00%*	*Annual probability of progression converted to rates and adjusted for cycle length, then re-converted to probability*	*(La Mattina et al., 2011)*
	Kidney-related death	*1.21%*	*Added mortality risk assumed in CKD stage 3*	*(Keith et al., 2004)*
CKD stage 4	*CKD stage 4*	*96.97%*	*1 - P(leaving state)*	*Calculation*
	CKD stage 5	*0.75%*	*Annual probability of progression converted to rates and adjusted for cycle length, then re-converted to probability*	*(La Mattina et al., 2011)*
	Kidney-related death	*2.28%*	*Added mortality risk assumed in CKD stage 4*	*(Keith et al., 2004)*
CKD stage 5	*CKD stage 5*	*95.43%*	*1 - P(leaving state)*	*Calculation*
	Kidney-related death	*4.57%*	*Added mortality risk assumed in CKD stage 5*	*(Keith et al., 2004)*
CKD stage 5 (after 14 cycles)	*CKD stage 5 (in cycle 14 only)*	*0.00%*	*1 - P(leaving state)*	*Calculation*
	No CKD via renal transplant (in cycle 14 only)	*99.33%*	*Patients transplanted between 32 - 35 mths on waiting list*	*(Organ Donation, 2014)*
	Operative Death	*0.7%*	*Operative death rate*	*(Aylin et al., 2013)*

Post-renal transplant ("No CKD [T]")	*No CKD (T)*	*98.25%*	*Move back into start of model*	*Assumption*
	CKD stages 1 to 2	*1.75%*	*Same progression chance as before transplant*	*(Abad et al., 2011)*
	Death (non-kidney related)	*0.00%*	*Natural mortality is assumed and is already accounted for in the hepatic model*	*(Organ Donation, 2014)*

**(Abad., 2011[123]) found that there were no differences between patients with first transplantation and those with a repeat transplantation in 419 renal transplant patients*

Source: Novartis Submission Table 52.

Relationship between the core hepatic model and the renal sub-model

Novartis report that *declining renal function can have a major impact on graft and patient survival.* and that in *transplant patients, chronic kidney disease is associated with an increased predisposition to cardiovascular events, an increased risk of hospitalisation, hepatic allograft dysfunction and an increase in mortality when compared with patients with preserved renal function (Ojo et al., 2003, McGuire et al., 2009, Charlton et al., 2009).*

However the company also reports that the hepatic core model and the renal sub-model are not linked in any way with the exception of death, which can occur as a result of progression in either model, and is considered an absorbing state in both models. Of note is that patients on the hepatic and renal sub-model will be on the same immunosuppressive regimen (i.e. EVR+rTAC or comparator treatments) and that these will impact hepatic and renal outcomes accordingly.

- ## Health related quality of life

As health-related quality of life (HRQoL) data were not collected as part of the H2304 trial, the company carried out a systematic literature review to identify health state utility values.

The ERG are satisfied that the primary searches as presented by the company in their initial submission meet the requirements to systematically identify relevant evidence on HRQoL and the information provided allows the replication of these methods.

In clarification, the company confirmed that a pragmatic literature search had been undertaken to populate their renal sub-model. These searches have not been reported, so the ERG cannot, therefore, validate or replicate the identification of studies.

The search strategy takes the following form and it was last updated in July 2014:

(Terms for liver or hepatic graft or transplantation) AND (terms for HRQL and various instruments, such as EQ5D)

The company searched all of the required resources for this section of the submission.

Novartis identified 7 studies which are reported in Table 57 of their submission.

Novartis report that even though the studies provided data for post-transplant patients, none of them reported disutility data for adverse events or utility data for acute or mild rejection.

Hepatic model

Novartis report the QoL values used in their hepatic model in Table 58 of their submission. The ERG found a reporting mistake in the table and given that the other values reported were amended by the company for use in the economic model, the ERG report the final values used by the company in the excel model in Table 26 below. Values in bold represent misreported estimates. This is further explored in Section 5.2.

The SPT state yields a 0.58 utility. Novartis state that patients have to undertake immunosuppressive therapy and that there is a loss in QoL associated with this kind of therapy.

This value was obtained by calculating an average of the utility values reported in Ratcliffe et al, 2002 (Table 27) for different time points weighted by the respective sample sizes. The values obtained from the QoL study were adjusted for informative dropout (in the Radcliff et al, 2002 analysis) and report a p-value of 0.55 for the test of difference in QoL values over time, which was therefore not statistically significant.

Novartis' clinical experts advised that the AR, ASRR and MCR state are asymptomatic. Therefore the company assumed that the QoL experienced in these health states is the same as the QoL experienced in the SPT state. Nonetheless Novartis report that patients in these states might require hospitalization.

Furthermore, it was assumed that pre-transplanted patients in a state of liver failure experience the same QoL as patients in the SCR (i.e. graft loss state).

Finally, patients in the SPT (via re-transplant) state were also assumed to experience the same QoL as patients in the initial SPT state.

Novartis report that the baseline time point was immediately (one day) post-transplant but that patients entered the model in the SPT state. Thus, *the baseline HRQL score was equalled to the score in the first cycle. However, disutility was applied upon entry if the patient experienced an adverse event upon entering the model.*

Table 26. Hepatic utilities used in the excel model

Health state	Utility value	Source
Stable post-transplant state	0.58	Novartis calculations and Radcliffe et al. (2002)
Acute rejection	0.58	Assumption: AR is asymptomatic hence utility is the same as for SPTA

Acute steroid resistant rejection	0.58	Assumption: ASSR is asymptomatic hence utility is the same as for SPTA
Mild chronic rejection	0.58	Assumption: MCR is asymptomatic hence utility is the same as for SPTA
Graft loss (severe chronic rejection)	**0.53**	**The manufacturer report a value of 0.43 for this states but the economic model uses 0.53.**
Stable PT via liver re-transplant	0.58	Assumption and Radcliffe et al. (2002)
Death	0	Assumption

Source: ERG table based on Novartis excel model

Table 27. Weighted average of QoL data – Novartis analysis to estimate utility in the SPT state

Time point	*Utility score adjusted for informative dropout*	*Sample size*	*Weighted average*
3 months post-transplantation	*0.57*	*309*	*0.5804*
6 months post-transplantation	*0.60*	*301*	
12 months post-transplantation	*0.59*	*329*	
24 months post-transplantation	*0.56*	*289*	

Source: ERG table based on Novartis excel model

HRQoL was assumed constant over time. Only the transition through health states drives changes in QoL of patients. Novartis explain that two of the QoL studies identified in the review showed no increase in HRQoL over time for transplanted patients, while 3 of them showed an improvement. However, it is stated that the level of graft rejection and graft functioning of the study patients is not reported *so it is difficult to establish why patients in some studies may improve over time and those in some other studies may not*.

Novartis add that *it would be incorrect to assume an improving trend is applicable to all patients, regardless of their state of rejection in the model as the published utility trend data is likely derived from a mixed cohort of patients including both rejecters and none rejecters.*

Renal sub-model

Novartis report the QoL values used in their renal model in Table 58 of their submission. The ERG reproduce the values in Table 28 below.

It was assumed that the no CKD state yields a utility value of 0.83. It is stated that patients in this state suffer from no detriment in their QoL, beyond that incurred in the hepatic model.

The main source for the CKD utility data is Neri et al, 2012. Novartis state that for CKD stage 1/2 patients observe a mildly reduced kidney function and that symptoms may include elevated blood pressure.

Patients in CKD stage 3 start to present complications such as hypertension, anaemia, kidney pain and early bone disease.

Patients in CKD stage 4 are reported to have severely reduced kidney function. Hypertension, anaemia, cardiovascular problems, nausea, insomnia and concentration problems are all common.

Patients in CKD stage 5 are reported to have very severe, or end-stage kidney failure and feel severely ill. Dialysis is required at this stage.

Finally, patients in the no CKD state (via re-transplant) are assumed to return to a reasonably normal HRQL, but will undergo immunosuppression. Novartis state that these patients experience a similar HRQoL to the SPT state in the main (hepatic rejection) model.

Table 28. Renal utilities used in the excel renal sub-model

Health state	Utility value	Source
No CKD	0.83	Assumption
CKD stages 1 to 2	0.64	Neri et al, 2012
CKD stage 3	0.58	Neri et al, 2012
CKD stage 4	0.49	Neri et al, 2012
CKD stage 5	0.28	Neri et al, 2012
No CKD via renal transplant	0.81	Liem et al, 2008
Death	0	Assumption

Source: ERG table based on Novartis excel model

Relationship between the core hepatic model and the renal sub-model

Novartis' clinical experts advised that the different states in the core hepatic model and renal sub-model could be considered independent in terms of impact in QoL.

Furthermore the company considered 3 different approaches for dealing with the potential double counting of utility losses in simultaneous health states (for example hepatic rejection and renal dysfunction). These were the multiplicative method, the additive method and the minimum method. The company used the minimum method in their base case analysis.

The minimum method assumes that the minimum utility value for simultaneously occurring health states is considered, for example, the utility for CKD stage 5 is 0.4 and utility for AR is 0.7. The joint state utility for those with both CKD stage 5 and acute rejection is 0.4.The additive model would produce instead a utility of 0.1 (=1-[1-0.4+1-0.7]). The multiplicative model instead would yield 0.28 (0.4 X 0.7).

More details on the other methods can be found on the company's submission page 177.

The company excluded from the analysis health effects associated with hyperacute rejection, HCV and HCC as it was considered that the choice of immunosuppressive regimen has no impact on these outcomes (thus there would not be differences between treatment arms).

Superseded
See Erratum

- Adverse events

Novartis estimated the occurrence of AE associated with the different treatment regimens in the economic model based on NMA data and standard product characteristics (SPCs) for the respective drugs' respective SPCs.

Besides the impact on renal functioning, other treatment related AEs considered were hypertension, diabetes mellitus, infections (bacterial, opportunistic, cytomegalovirus - CMV and fungal), tremor and insomnia.

The recurrence of HCV and HCC was not considered in the model as Novartis state that clinical literature and consultation with clinical experts indicated that there is no effect of the immunosuppressive regimen on these variables.

AE data are reproduced in Table 29 for the different treatment arms. Table 30 and Table 31 report the disutility values and the costs associated with the different AEs, respectively. Costs were reported to have been inflated to 2013 prices where necessary.

The systematic literature searches undertaken by the company did not identify any relevant QoL data associated with specific immunosuppressive-related AEs. Therefore the company used Sulivan et al, 2011 to obtain QoL data.

Table 29. Incidence of AEs in the economic model

AE	EVR+rTAC	AZA+TAC	MMF+TAC
Hypertension	40.3%	57.8%	23.9%
New onset diabetes	15.7%	18.3%	11.6%
Infection	65.7%	60.3%	62.6%
Herpes	0.4%	5.9%	10.1%
Tremor	10.2%	35.5%	33.9%

Insomnia	6.9%	47.0%	52.3%
CMV	4.1%	8.0%	5.8%

Table 30. AEs disutility data

AE	Disutility value	Source	Justification
Hypertension	-0.04	(Sullivan et al., 20111)	No specific data on hypertension in post-liver transplant patients was identified ; general disutility associated with hypertension was used (098 Essential Hypertension).
New onset diabetes	-0.06	(Sullivan et al., 20111)	No specific data on new onset diabetes in post-liver transplant patients was identified ; general disutility associated with diabetes was used (049 Diabetes mellitus without complications).
Infection	-0.01	Assumption	No data relating to infection disutility was identified.
Herpes	-0.001	(Sullivan et al., 20111)	No specific data on herpes in post-liver transplant patients was identified ; general disutility associated with herpes was used (200 other skin disorders).
Tremor	-0.003	(Sullivan et al., 20111)	No specific data on tremor in post-liver transplant patients was identified ; general disutility associated with alcohol related disorders was used (alcohol-related disorders).
Insomnia	-0.04	(Sullivan et al., 20111)	No specific data on insomnia in post-liver transplant patients was identified ; general disutility associated with sleep disorders was used.
CMV (cytomegalovirus)	-0.01	Assumption	No data on CMV disutility was identified; assumed the same as infection.

Table 31. AEs cost data

Adverse event	Items	Treatment cost	Reference in submission
Hypertension	Single treatment cost	£769	(Department of Health, 2013)
New onset diabetes	Single treatment cost	£1,840	(Davidson *et al.*, 2003)
Infection	Single treatment cost	£649	(Department of Health, 2013) (PA18B) Minor Infections with CC Score 0
Herpes	GP visit	£55	(Curtis, 2013)
Tremor	GP visit	£55	(Curtis, 2013)
Insomnia	GP visit	£55	(Curtis, 2013)
CMV (cytomegalovirus)	Single cost includes IV Ganciclovir IV, testing, imaging, hospital stay (14 days)	£11,365	(Annemans *et al.*, 2002)

- ## Resources and costs

The company carried out a systematic literature review to identify resource use data. Three studies were selected for data extraction (Foxton et al, 2010, Anneman et al, 2002 and Burroughs et al, 1992)

The ERG are happy to accept the primary searches as presented by the company in their initial submission.

The search strategy takes the following form and it was last updated in August 2014, so the results are considered current for this submission:

(Terms for liver or hepatic graft or transplantation) AND (terms for costs or economics)

The company searched all of the required resources for this section of the submission. The only limitation of the search strategy was that, following clinical advice, the company used a date limit on these searches of 1990-Current.

The model submitted by Novartis used costs based on the NHS & PSS perspective. Costs included in the model are drug costs and disease management costs (such as monitoring costs and outpatient visits).

Estimates of resource use for the hepatic model were obtained from the University Hospitals Birmingham Foundation Trust and validated by Novartis' clinical advisors.

The model heath states are claimed to have been costed to represent English clinical practice. The costs considered for the hepatic health state are presented in Table 32, whilst the costs used in the renal health states are presented in Table 33.

Other miscellaneous costs were not considered in the economic analysis.

Table 32. Health state costs used in the hepatic model

Health state	Resource	Resource use	Unit cost	Source
Stable post-transplant (SPT)	GP Visit	1 visit	£55.00	Resource use – assumption Unit costs - (Curtis, 2013)
	Liver function test	1 test	£5.43	Resource use – UHBFT, 2012 Unit costs - (NICE, 2012), inflated to 2013
	Full blood count	1 test	£3.32 per test	Resource use – UHBFT, 2012 Unit costs- (NICE, 2012), inflated to 2013
	Urine and electrolytes	1 test	£8.44 per test	Resource use – UHBFT, 2012 Unit costs- (NICE, 2012), inflated

				to 2013
	Total health state cost per cycle		**£72.19**	
Acute rejection (AR)	Hospital inpatient stay	3 days	£337.94 per day	Resource use – assumption Unit costs - Department of Health, 2013 (GC01D) Liver Failure Disorders with single intervention gastroenterology
	Hospital outpatient visits	1 days	£140.73	Resource use – assumption Unit costs - University Hospitals Birmingham Foundation Trust, 2013
	GP visit	3 visits	£55.00	Resource use – assumption Unit costs - Curtis, 2013
	Liver function test	3 tests	£5.43 per test	Resource use – UHBFT, 2012 Unit costs - (NICE, 2012), inflated to 2013
	Full blood count	3 tests	£3.32 per test	Resource use – UHBFT, 2012 Unit costs - (NICE, 2012), inflated to 2013
	Urine and electrolytes	3 tests	£8.44 per test	Resource use – UHBFT, 2012 Unit costs - (NICE, 2012), inflated to 2013
	High dose steroid therapy (methylprednisolone)	Average 1.14 courses/patient	£32.16 per course	Resource use – assumption Unit costs - Commercial Medicines Unit, 2014
	Liver biopsy	1 test	£514.63	Resource use – assumption Unit costs - Department of Health, 2013 (FZ61Z)
	Total health state cost per cycle		**£1,922.32**	
Acute steroid-resistant rejection (ASRR)	Hospital inpatient stay	14 days	£337.94 per day	Resource use – assumption Unit costs - Department of Health, 2013 (GC01D)
	Hospital outpatient visits	1 days	£140.73	Resource use – assumption Unit costs - University Hospitals Birmingham Foundation Trust, 2013
	GP visit	6 visits	£55.00	Resource use – assumption Unit costs - Curtis, 2013
	Liver function test	5 tests	£5.43 per test	Resource use – UHBFT, 2012 Unit costs - (NICE, 2012), inflated to 2013
	Full blood count	5 tests	£3.32 per test	Resource use – UHBFT, 2012

				Unit costs - (NICE, 2012), inflated to 2013
	Urine and electrolytes	5 tests	£8.44 per test	Resource use – UHBFT, 2012 Unit costs - (NICE, 2012), inflated to 2013
	Antilymphocyte therapy	14 days	£703.99 per 14 day course	Resource use – assumption Unit costs - Commercial Medicines Unit, 2014
	Liver biopsy	1 test	£514.63	Resource use – assumption Unit costs - Department of Health, 2013 (FZ61Z)
	Total health state cost per cycle		**£6,506.47**	
Mild chronic rejection (MCR)	Hospital outpatient visits	3 days	£140.73	Resource use – assumption Unit costs - University Hospitals Birmingham Foundation Trust, 2013
	GP visit	3 visits	£55.00	Resource use – assumption Unit costs - Curtis, 2013
	Liver function test	3 tests	£5.43 per test	Resource use – UHBFT, 2012 Unit costs - (NICE, 2012), inflated to 2013
	Full blood count	3 tests	£3.32 per test	Resource use – UHBFT, 2012 Unit costs - (NICE, 2012), inflated to 2013
	Urine and electrolytes	3 tests	£8.44 per test	Resource use – UHBFT, 2012 Unit costs - (NICE, 2012), inflated to 2013
	Total health state cost per cycle		**£638.77**	
Severe chronic rejection (graft loss) (SCR)	Hospital inpatient stay	3 days	£337.94 per day	Resource use – assumption Unit costs - Department of Health, 2013 (GC01D)
	Hospital outpatient visits	14 days	£140.73	(University Hospitals Birmingham Foundation Trust, 2013158)
	GP visit	3 visits	£55.00	Resource use – assumption Unit costs - Curtis, 2013
	Liver function test	6 tests	£5.43 per test	Resource use – UHBFT, 2012 Unit costs - (NICE, 2012), inflated to 2013
	Full blood count	6 tests	£3.32 per test	Resource use – UHBFT, 2012 Unit costs - (NICE, 2012), inflated to 2013
	Urine and	6 tests	£8.44 per test	Resource use – UHBFT, 2012

	electrolytes			Unit costs - (NICE, 2012), inflated to 2013
	Total health state cost per cycle		**£2,126.36**	
Stable post-transplant (first cycle after re-transplant) (SPT post-transplant)	GP visit	1 visit	£55.00	Resource use – assumption Unit costs - Curtis, 2013
	Liver function test	1 test	£5.43 per test	Resource use – UHBFT, 2012 Unit costs - (NICE, 2012), inflated to 2013
	Full blood count	1 test	£3.32 per test	Resource use – UHBFT, 2012 Unit costs - (NICE, 2012), inflated to 2013
	Urine and electrolytes	1 test	£8.44 per test	Resource use – UHBFT, 2012 Unit costs - (NICE, 2012), inflated to 2013
	Liver transplant	1 transplant	£20,358.60	Resource use – assumption Unit costs - Department of Health, 2013 (GC01D)
	Total health state cost per cycle		**£20,430.80**	

For the renal sub-model, Novartis reported NICE CG182 and Kerr et al, 2012 as the two sources used to obtain resource use data.

Finally, all of the costs presented for the CKD health states were wrong in the submission (when compared with the economic model). This was is a misreporting problem, as the resource use and the unit costs in the submission are correct and match the ones in the excel model, used to derive the total costs. We present the correct estimates in Table 33.

Table 33. Health state costs used in the renal sub-model

Health state	Resource	Resource use	Unit cost	Source
No CKD	No costs			
CKD stages 1 to 2	GP visit	0.5 visits	£55.00	Resource use – Kerr et al, 2012 Unit costs - Curtis, 2013
	eGFR test	0.25 tests	£150.00 per test	Resource use – NICE, 2014 Unit costs - NICE, 2014
	Total health state cost per cycle		**£65***	
CKD stage 3	GP visit	0.5 visits	£55.00	Resource use – Kerr et al, 2012 Unit costs - Curtis, 2013
	eGFR test	0.50 tests	£150.00 per test	Resource use – NICE, 2014

				Unit costs - NICE, 2014
	Total health state cost per cycle		**£103***	
CKD stage 4	Hospital outpatient visits	0.25 days	£140.73	(University Hospitals Birmingham Foundation Trust, 2013158)
	eGFR test	1 test	£150.00 per test	Resource use – NICE, 2014 Unit costs - NICE, 2014
	Renal ultrasound	0.25 tests	£52.00 per scan	Resource use – NICE, 2014 Unit costs - Department of Health, 201. Ultrasound Scan, less than 20 minutes – (RA23Z)
	Total health state cost per cycle		**£123***	
CKD stage 5	Hospital outpatient visits	0.25 days	£140.73	Resource use – NICE, 2014 Unit costs - University Hospitals Birmingham Foundation Trust, 2013
	GP visit	0.5 visits	£55.00	Resource use – Kerr et al, 2012 Unit costs - Curtis, 2013
	eGFR test	1.0 tests	£150.00 per test	Resource use – NICE, 2014 Unit costs - NICE, 2014
	Haemodialysis	0.25 of annual course	£30,020.01 per year	Resource use – NICE, 2014 Unit costs- Kerr et al., 2012 Inflated to 2013 prices from 2009
	Peritoneal dialysis	0.25 of annual course	£25,069.32 per year	Resource use – NICE, 2014 Unit costs- Kerr et al., 2012 Inflated to 2013 prices from 2009
	Waiting list clinic attendances	0.25 of annual cost	£3,709.58 per patient year	Resource use – Kerr et al, 2012 Unit costs- Kerr et al., 2012 Inflated to 2013 prices from 2009
	Total health state cost per cycle		**£14,912***	
Post renal transplant	Renal transplant	1 transplant	£17,266.00 per transplant	Resource use – NICE, 2014 Unit costs- Department of Health, 2013 (LA01A)
	Post-transplant on going care costs	0.25 of annual cost	£9,137.23 per patient year	Resource use – Kerr et al, 2012

				Unit costs- Kerr et al., 2012 Inflated to 2013 prices from 2009
	Total health state cost per cycle		**£2,289* (plus one-off cost of renal transplant)**	

* Values corrected by the ERG

Intervention and comparators' costs

Novartis report that 3 components were considered for the estimation of the intervention and comparator costs:

- The list price of the drug regimen
- Costs relating to the administration of the regimens.
- Costs related to the monitoring of patients on the regimens.

The price of individual technologies is reported to have been taken from the British National Formulary (BNF, 2014) and the Commercial Medicines Unit (Commercial Medicines Unit, 2014161) websites (searched July 2014), with the exception of EVR, the cost of which was directly provided by the company. Novartis have used the Prograf brand price to cost the use of TAC in the model.

All therapies were assumed to be taken orally. The only exception is intravenous MMF which is administered in the first 2 weeks post-transplant. Novartis assumed that patients are still in the hospital at this point as they are recovering from their transplant thus no additional administration costs were considered (in-patient stay costs are considered to capture these costs).

No on-going monitoring costs were included as it was assumed that that these were included in the consultations costs.

Table 34 shows the cost per milligram for each technology within each regimen, the average daily cost of each drug within each regimen and the average annual cost of the regimen, both taking into account the average body weight of 73.9kg of a patient, for the doses that are determined by kilogram of body weight. Patient weight was simulated in the model.

Table 34. Intervention and comparators' costs

Regimen/drug	Cost per mg	Source	Average daily cost	Average annual cost
EVR+rTAC				**£9,084.89 per year**
Everolimus	£9.90	Company internal	£18.15 /day	
Tacrolimus	£1.61 per Kg	Prograf® (BNF, 2014[127])	£6.68 /day	
Corticosteroid	£0.01	Prednisolone (Commercial Medicines Unit, 2014[161])	£0.06 /day	
AZA+TAC				**£3,967.22 per year**
Azathioprine	£0.001 per Kg	non-proprietary (Commercial Medicines Unit, 2014[161])	£0.08 /day	
Tacrolimus	£1.61 per Kg	Prograf® (BNF, 2014[127])	£10.73 /day	
Corticosteroid	£0.01	Prednisolone (Commercial Medicines Unit, 2014[161])	£0.06 /day	
MMF+TAC				**£4,654.56 per year**
MMF intravenous	£0.02	CellCept® (BNF, 2014[127])	£1.40 /day	
MMF oral	£0.0004	non-proprietary (Commercial Medicines Unit, 2014[161])	£1.11 /day	
Tacrolimus	£1.61 per Kg	Prograf® (BNF, 2014[127])	£10.73 /day	
Corticosteroid	£0.01	Prednisolone (Commercial Medicines Unit, 2014[161])	£0.06 /day	

- ### Discounting

All costs and health benefits were discounted at a 3.5% rate as recommended by NICE.

- ### Sensitivity analysis

Probabilistic sensitivity analysis (PSA) was undertaken by Novartis. Distributions used to run PSA were also reported in the submission (Table 67). Deterministic sensitivity analysis was not run as per the company statement.

Scenario analysis was undertaken to examine the impact of changing the structure of the economic model on the model outputs. The outputs of this analysis are reported in Section 5.3 of this report.

- ### Model validation

It is stated by Novartis that the model was checked for internal quality with regards to formulae, programming string and data inputs, however the ERG discovered several important logic errors and formulae in the economic model sent to us by Novartis. Data inputs also presented inconsistencies when compared with values reported in the submission.

Model outputs have not been externally validated. This means that life expectancy, graft loss and renal function weren't compared with any available literature. Whenever possible, the ERG tried to validate model main outcomes with external sources.

It is also mentioned in the submission that expert opinion was sought by Novartis to validate several model inputs and assumptions.

5.2 Critique of approach used

In this section, we comment on Novartis approach and methodology. First, we consider the model against checklists of good practice. Then we critically appraise the model structure and data as well as the methods used in the cost effectiveness analysis.

5.2.1 Critical appraisal frameworks

Novartis' economic analysis was assessed against three widely used study quality checklists for economic models:

- NICE Reference Case (NICE, 2008).
- Drummond assessment criteria (Drummond et al., 1997).
- Criteria for decision model-based economic evaluations (Philips et al., 2006).

Table 35. Critical appraisal checklist based on NICE Reference Case (NICE, 2008)

NICE reference case requirement		Critical appraisal	Reviewer comment
Defining the decision problem	The scope developed by NICE	✓	
Comparator	Therapies routinely used in the NHS, including technologies regarded as current best practice	?	The ERG believe that reduced TAC monotherapy should have been included as a comparator. since this is widely used in clinical practice; Furthermore, AZA and MMF given with standard doses of TAC do not necessarily reflect clinical practice at later stages of immunosuppressive therapy. These drugs should have been considered with concomitant doses of reduced TAC.
Perspective on costs	NHS and PSS	✓	NHS & PSS
Perspective on outcomes	All health effects on individuals	✓	
Type of economic evaluation	Cost-effectiveness analysis	✓	
Synthesis of evidence on outcomes	Based on a systematic review	?	Based primarily on single trial (H2304) evidence and on NMA which validity is unclear
Measure of health benefits	QALYs	✓	
Source of data for measurement of HRQL	Reported directly by patients and/or carers	✓	EQ- 5D survey
Source of preference data for valuation of changes in HRQL	Representative sample of the public	✓	EQ- 5D survey
Discount rate	3.5% pa for costs and health effects	✓	
Equity weighting	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	✓	

Note: ✓ indicates 'clear'; X indicates 'concerns'; ? indicates 'some concerns'.

Table 36. Critical appraisal checklist from Drummond and colleagues (Drummond et al. 1997)

Item	Critical appraisal	Reviewer comment
Has the correct patient group/population of interest been clearly stated?	?	There are some differences between the trial population and the typically presenting UK population. Furthermore there is considerable heterogeneity across study populations in the NMA.
Is the correct comparator used?	?	The ERG believe that reduced TAC monotherapy should have been included as a comparator. since this is widely used in clinical practice; Furthermore, AZA and MMF given with standard doses of TAC do not necessarily reflect clinical practice at later stages of immunosuppressive therapy. These drugs should have been considered with concomitant doses of reduced TAC.
Is the study type reasonable?	?	A patient simulation model was used. The ERG are not convinced this is necessary.
Is the perspective of the analysis clearly stated?	✓	UK NHS PSS
Is the perspective employed appropriate?	✓	NHS Reference Costs
Is the effectiveness of the intervention established?	?	Quality of H2304 is good in establishing the renal sparing effect of EVR+rTAC. The effectiveness absolute values obtained for the NMA rely on a non-robust analysis.
Has a lifetime horizon been used for analysis, if not has a shorter time horizon been justified?	x	The model ran for 80 years. The time frame seems unnecessarily high, adding to the computational burden of the model. After 80 years in the model, these patients will be, on average, 134 year old.
Are the costs and consequences consistent with the perspective employed?	✓	All costs are presented from the UK NHS & PSS perspective
Is differential timing considered?	✓	All future costs and benefits are discounted with a 3.5% rate.
Is incremental analysis performed?	✓	
Is sensitivity analysis undertaken and presented clearly?	?	Probabilistic sensitivity analyses is reported but it lacks robustness.

Note: ✓ indicates 'clear'; X indicates 'concerns'; ? indicates 'some concerns'.

Superseded
See Erratum

Table 37. Critical appraisal checklist of Philips et al (2004) for model-based analysis

Dimension of quality		Critical appraisal	Comments
Structure			
S1	Statement of decision problem/objective	✓	The submission from Novartis considered the use of everolimus (Certican®) in combination with reduced dose tacrolimus with corticosteroids for the initial 6 months of therapy, for adults on maintenance immunosuppressive therapy following allogeneic liver transplantation.
S2	Statement of scope/perspective	✓	NHS & PSS perspective was implemented. Cost and benefit inputs were consistent with this. Scope of the model stated.
S3	Rationale for structure	?	The model would have benefited from further exploring the relationship between hepatic model and renal sub-model.
S4	Structural assumptions	X	Generally, the ERG are not convinced by some of the structural assumptions. These are explored in Section 5 of this report.
S5	Strategies / comparators	?	It is not clear to the ERG why MMF and AZA weren't considered with reduced doses of TAC and why TAC monotherapy wasn't included as a comparator.
S6	Model type	X	A patient simulation model was used. The ERG are not convinced this is necessary and Novartis did not justify this decision adequately.
S7	Time horizon	X	The model ran for 80 years. The time frame seems unnecessarily high, adding to the computational burden of the model. After 80 years in the model, these patients will be, on average, 134 year old.
S8	Disease states / pathways	?	The ERG are not convinced of the clinical plausibility of the MCR state.
S9	Cycle length	?	Cycle length is 3 months .Clinical opinion sought by the ERG indicated that shorter cycles would capture events and outcomes more accurately.
Data			
D1	Data identification	?	The effectiveness of EVR+rTAC was well sourced in H2304. The comparative effectiveness of MMF+TAC and AZA+TA relied on the NMA which lacked clarity and robustness.
D2	Pre-model data analysis	?	No information given

Dimension of quality		Critical appraisal	Comments
D2a	Baseline data	?	Baseline data used in the simulation model presented similar characteristics across H2304 and comparator studies. However, other potential treatment modifiers presented different baseline values across studies.
D2b	Treatment effects	X	The ERG do not feel confident in the NMA results.
D2c	Quality of life weights (utilities)	?	HRQoL was not recorded in H2304. The ERG found some problems with the utility estimates used.
D3	Data incorporation	?	Data inputted in the model is generally poorly referenced in the submission.
D4	Assessment of uncertainty	?	A PSA is presented but the results lack robustness.
D4a	Methodological	?	Results reported lack face validity.
D4b	Structural	?	Results reported lack face validity.
D4c	Heterogeneity	?	Subgroup analysis was not undertaken in the model. NMA scenario analysis wasn't properly reported.
D4d	Parameter	?	Deterministic sensitivity analysis wasn't performed.
Consistency			
C1	Internal consistency	X	Even though Novartis claim to have sought validation for the excel model, the model contained several serious logical errors and the
C2	External consistency	?	Only expert opinion was sought for external validation.

Note: ✓ indicates 'clear'; X indicates 'concerns'; ? indicates 'some concerns'.

5.2.2 Critique of the modelling approach and structure

Novartis's cost-effectiveness model was developed as a patient simulation model and includes a core hepatic model and a renal sub-model.

The ERG focus on three specific issues in this subsection. These are the following:

1. Novartis decision to use a patient simulation modelling approach
2. Novartis decision to model a core hepatic model and a renal sub-model
3. Specific issues identified in Novartis model structure

Whilst the two first topics discuss the appropriateness of the modelling approach from a methodological point of view, the third one focus on specific issues identified in the company model.

1. Patient simulation model

It is worth specifying that the model developed by the company is a patient simulation state-transition model (as opposed to, for example a discrete event simulation model, which also falls under the broader characterization of a patient simulation model).

State-transition models consist of a discrete set of mutually exclusive health states which are evaluated at regular intervals (model cycles). When applied at the patient level (i.e. within a patient simulation framework) these are evaluated stochastically using samples drawn from statistical distributions to determine whether an individual patient experiences a particular transition given the probability of that transition occurring in that particularly cycle.

It is the ERG opinion that Novartis did not provide enough evidence to justify their methodological approach, nor did the company provide clear details regarding the approach and assumptions used.

Superseded
See Erratum

Novartis state that the use of a patient simulation model is appropriate when the patient flow is determined by time since last event or history of previous event (David et al, 2014). However there is no clear explanation as to why this argument applies to their analysis. Time dependency in the model only exists in relation to transplantation, which happens 30 days before the model begins. This means that this is a “fixed event” in time within the model timeline.

In fact, the patient flow in the economic model is only affected by time since last event in the severe chronic (SCR) rejection state. In the renal sub-model transition probabilities are not time dependant and also do not depend on history of previous events.

The ERG are not convinced that a patient simulation model is necessary to deal with the time dependency of transition probabilities in terms of occupation of subsequent health states in this case as:

- The change in transition probabilities over time from the SPT to the AR state (Table 22) is only dependant on time since transplantation. As time since transplantation progresses simultaneously with the succession of cycles in the model, the change in the transition matrix would be straightforward to implement in a cohort state-transition model. Furthermore, the model considers 3-month cycles and after 12 months all transition probabilities are assumed constant in the model. The same argument applies to transition to the MCR state.

- Episodes of AR are assumed to occur with the same probability, regardless of the number of AR episodes previously experienced by the patient, making the previous history of AR events irrelevant.

- Patients are assumed to stay in the SCR state for 2 cycles as they are assumed to be re-transplanted only after the end of six months. Mortality rates in the SRS state also change from the first cycle spent in the state to the second one. However, this would be easily resolved by creating a tunnel state in a traditional cohort approach. Dividing the

SCR state into SCR1 and SCR2 for example, would allow different transitions probabilities to be associated with each state.

- In the case of a re-transplantation, transition probabilities from the stable state onwards are the same as the ones used in the case of the first transplant.

The company also state that a patient level simulation approach facilitates *between-model calculations* involving the hepatic core model and the renal sub-model. This argument is irrelevant in the case of Novartis analysis as there are no interactions between events in the two models (as the company acknowledges). The only link between these models is that calculations in the excel model were developed to account for the fact that if one patient dies in one model he is recognized to be dead in the other. Again, this is something which could be done in a cohort structure.

Another reason why patient level models might be appropriate is that such models allow to consider individual-level patient heterogeneity. If there are clinical variables which differ between patients (for example age) and which are known to have a non-linear relationship with the model outcomes (e.g. costs and QALYs), then estimating the model outcomes for a cohort of patients using average characteristics (e.g. mean age at starting treatment) will provide a biased estimate of the average outcome across the population to be treated (Davis et al, 2014).

However, in situations where such clinical factors can be identified prior to treatment initiation, then a cohort approach can be followed which includes subgroup analysis to account for the average outcomes in these specific groups of patients (provided the outcomes within those subgroups are expected to be reasonably homogeneous) and the results from the subgroups averaged with those for the rest of the population using weights according to their relative sizes. (Davis et al, 2014).

Novartis do not mention any justification of their methods in terms of the appropriateness or the relevance of their patient simulation approach in terms of dealing with patient heterogeneity. However, in their excel model they have simulated a group of patients with varied baseline characteristics in terms of the clinical parameters presented in Table 38 by randomly sampling values from population distributions. They use the mean eGFR baseline level to determine the starting CKD level in the renal sub-model (as explained in Section 5.1 above) and the baseline body weight to determine drug doses and costs in the hepatic model. Baseline age and gender only have an impact on the natural mortality rate.

Analysing the parameters in Table 38 against the hypothesis of these clinical variables having a non-linear relationship with the model outcomes as per the DSU technical document (Davis et al, 2014) it seems unlikely that this is the case. It seems reasonable to assume that body weight has a linear relationship with costs, as the higher the body weight is for the patient, the higher the drug costs will be (as more quantity is needed to achieve the same trough levels). The same applies for eGFR where a lower baseline value always leads to a higher risk of CKD progression.

Table 38. Baseline caracteristics varied in the patient simulation model

Variable	*Value*	*CI (distribution)*	*Reference to section in submission*
Baseline age	*53.8 years*	*18.0 to 70.0 range (Lopez et al., 201340) (Normal distribution)*	*Trial results section 6.5*
Baseline gender	*50%*	*45% to 55% (Normal distribution)*	*Trial results section 6.5*
Baseline body weight (kg)	*82.6*	*Range 59.1 to 89.7 (Standard deviation 14.82) (Lopez et al., 201340) (Normal distribution)*	*Trial results section 6.5*
Baseline mean eGFR immediately post-liver transplant	*81 mL/min/1.73 m2*	*17.5 to 259.1 range (Lopez et al., 201340) (Normal distribution)*	*Trial results section 6.5*

Interestingly, the company reports potential treatment effect modifiers across the NMA studies, identified in discussions with clinical experts before undertaking their analysis. These were considered to be donor age, cold ischaemic time (organ quality), HCC, exclusion of concomitant steroids in the first 3 months, diabetes and hypertension. BMI is also mentioned to be a treatment modifier. From this list, only body weight (which is related with body weight) was included in the economic model.

Further to this, the company undertook subgroup analysis in their H2304 trial as reported in Section 4.2. The subgroups considered were age, gender, race, eGFR, HCV status, MELD score and cause of ESLD for the composite clinical endpoint (tBPAR, graft loss or death), while age, gender, race, region, hepatitis C serology status, eGFR strata, MELD categories and ESLD were the subgroups considered for the renal outcomes. **The statistical analysis undertaken showed no statistically significant relationships between the subgroup factors and both effectiveness outcomes**.

Based on assessment of patient heterogeneity, and the patient baseline characteristics simulated in the economic model, it seems that there was no need for a patient simulation model, and that a cohort approach would have sufficed.

In conclusion, it is not clear to the ERG why a patient simulation state-transition model is required in this situation. The company has not reported sufficient details or a solid justification for their approach. It seems that a traditional cohort state-transition model would be appropriate in this case.

One apparent disadvantage (albeit not a significant one) of using a cohort approach would be that it would have required the addition of one extra health state (the tunnel state for the SCR state).

The disadvantages of patient-level state transition models are:

- Patient-level state-transition models built in excel require great care as the logic required to specify the transition matrices may be complex and difficult to check for errors (Davis e al, 2014).

- PSA is more difficult to run. Due to its computational burden, Novartis ran 5,000 simulations in their PSA (compared with the 10,000 simulations used in the base case). This is a problem as it affects the reliability of results (this is further explored in the next section).

- It is considerably more difficulty to understand how sensitive the model outcomes are to changes in the model parameters or assumptions. On one hand there is a greater computational time requirement (it took the ERG around 1 hour to run the model every time we changed an input value) and on the other hand given that the model outcomes are generated by randomly sampling parameters it is not possible to replicate the random number streams and thus isolate the impact of changing one particular aspect of the model.

Superseded
See Erratum

2. Core hepatic model and renal sub-model

Novartis explain that the decision to capture the renal sparing effect associated with EVR+rTAC through a renal sub-model *was considered important because the treatment effect* [survival of liver graft] *has an impact on more than one aspect of a patients' health.* Additionally it is mentioned that *ISPOR good research practice guidelines encourage the design option of using sub-models to simplify the model structure (Karnon et al., 2012).*

Firstly, the ISPOR guidance cited by the company refers to discrete event simulation (DES) models. These are complex models, and even though they fall under the patient simulation model umbrella, cannot be compared with the company's patient simulation state-transition model from a methodological and technical complexity point of view. Therefore, the argument that sub-models simplify the model structure doesn't really apply to the company's model.

Secondly, and more importantly, the company reports that declining renal function can have a major impact on graft and patient survival, and the predisposition to cardiovascular events, as well as result in an increased risk of hospitalisation, hepatic allograft dysfunction and mortality.

Thirdly, Novartis state several times that while the core hepatic model attempts to track patients post-transplantation, the focus of this submission is to demonstrate the renal sparing effect of EVR+rTAC. This is in line with the H2304 clinical trial, where the statistical hypothesis under study were to test the non-inferiority of EVR+rTAC compared to TAC in terms of preventing acute rejection and organ loss and to test the superiority of EVR+rTAC compared to TAC in terms of preserving renal functioning.

Therefore the ERG believe that more emphasis should have been placed on the renal component of the economic model but also that more interaction between the 2 models should have been considered (perhaps within one broader model structure) as it is known that immunosuppressive therapy for liver post-transplantation has an impact not only on renal functioning but also, equally important, that renal functioning has an impact on graft survival.

3. Specific issues identified in the model

The ERG find that the reporting of the model structure and its assumptions lacked clarity. Also few justifications were provided as to why those assumptions were considered necessary and appropriate.

Superseded

In this subsection we discuss the different health states included in the economic model, the errors in formulae used to allocate patients into different health states, the cycle length of the model, the time horizon used and finally the number of simulated patients.

See Erratum

Health states in the economic model

The company's model structure is presented in Figure 8 above. The ERG found some inconsistencies in the representation of the model structure and the description of health states.

The hepatic core model describes 6 health states:

Stable post-transplant state (SPT):

Even though the company claims that patients enter the model in the SPT state 4 weeks after transplantation, this is not entirely accurate. Patients in the comparator arms of the model are assumed to enter the model immediately after transplantation, and incur the respective costs.

During the clarification process, the ERG asked the company to clarify at what point after transplantation the model begins. The company clarified that the model starts immediately (one day) post-transplant, as the use of comparator therapies begins at this time point.

However, for the EVR+rTAC arm of the model, patients only start receiving treatment 30 days after entering the model (i.e. 30 days post-transplantation). Costs and benefits for the EVR+rTAC arm were adjusted to reflect the later starting point accounting for 2 of the 3 months in the first cycle. The ERG found a mistake with regards to the AEs considered in the first cycle of the model. This s explored in the AEs subsection.

Acute rejection (AR):

Patients in the SPT state, suffering from AR (diagnosed by biopsy) move to this state, where they are given a high-dose steroid course of treatment for a minimum of 3 days and a maximum of 2 courses (i.e. 6 days of steroid treatment). If the treatment is successful, patients return to the stable SPT state. If treatment fails then patients move the ASRR state.

Of note is that the company mentions a maximum of 3 courses of steroid treatment. Nonetheless, a maximum of 2 courses was considered in the base case economic analysis.

Acute steroid resistant rejection (ASRR):

If steroid treatment fails, patients move to the ASRR state. In this state patients will undergo anti-lymphocyte therapy for 14 days. If the treatment is successful then patients return to the SPT state. If treatment is unsuccessful, patients move to the SCR state. Despite the fact that arrows in Figure 8 seem to indicate that patients can remain in this state for longer than 1 cycle, this is not the case as per the transition probability matrix used in the excel model.

Superseded
See Erratum

Severe chronic rejection (SCR):

If patients fail both steroid and anti-lymphocyte treatments for acute rejection, they will move into the SCR state. It is assumed that all patients entering the SCR state will suffer graft failure and will require a re-transplant.

Patients are assumed to stay in the SCR state for 2 cycles as they are assumed to be re-transplanted only after the end of six months (so only after being 2 cycles in this state).

Under the current structure, re-transplant patients go straight back to the SPT state, but EVR patients are only meant to enter this state 4 weeks after having their transplant. Therefore the model is not capturing this 4-week period during which events like hyperacute rejection and any other early rejections may happen. This could potentially bias the results as the outcomes captured for re-transplantation patients are better than they would be in real life. Nonetheless, this doesn't impact the economic base case analysis as virtually no patients are re-transplanted in the model across all treatment arms. Furthermore, as EVR is not expected to have an impact on the prevention of hyperacute rejection, the marginal difference across model arms could potentially even out, provided AZA and MMF also do not have an impact on this outcome.

Mild chronic rejection (MCR):

Patients are reported to be able to develop MCR from any state except SR and death. However, that is not the case, as in the excel model patients are only allowed to transition to the MCR state from the STP state. This is an asymptomatic state and it has been assumed that patients can only move to this state 1 year post-transplant. Moreover, due to some issues found in the excel formulae, patients can only progress to this state after they have completed 1 year in the STP state. This is discussed in the next section of the report.

The company also mention that patients can go from this state to the STP (Table 45 in the submission), however that is not the case in the base case economic analysis, where patients reaching the MCR state don't leave the state unless dead. This means that patients can stay in this state for years.

Clinical opinion sought by the ERG revealed that the state of mild chronic rejection seems somewhat vague and unfamiliar. Our clinical expert advisor did not see any valid (or justifiable) reason for patients only to progress to such state 1 year after transplantation. In fact the company recognizes that it is not clear what causes this state and that it is an asymptomatic one. The ERG question the clinical plausibility of this state and the plausibility of patients remaining in the state for years. The relevance of including this health state in the model is not clear. There is no change in the utility of life of patients when in this state, however the resource use associated with MCR is similar to that of the AR state. This is further explored in the next section.

<u>Hepatic-graft related death (HD):</u>

Superseded
See Erratum

Patients were assumed to have a higher mortality rate in the SCR state of the model. As patients are virtually kept on this state for 2 cycles, a 3% mortality rate is used during the 3 initial months of the patient being in the SR state to reflect the mortality rate of patients dying while on the waiting list. After the 6 months, when the patient will undergo surgery, a 0.7% operative death rate is applied to reflect patients dying during surgery.

The ERG found some mistakes in the formulae related to the mortality rates. This will be further explored in Section 5.2.3.

The health states captured in the renal model were taken from the NICE guidance on CKD progression and are presented in Table 21. The model structure seems to accurately capture the evolution of CKD disease. However in the context of the decision problem faced in this submission, the ERG's opinion is that a more interactive model linking hepatic and renal outcomes would be better suited.

Table 25 shows the increase in mortality associated with the different stages of CKD. It is appropriate to account for the increase in mortality as renal disease progresses.

A background natural mortality rate was also used for both hepatic and renal models, which is appropriate.

<u>Allocation of patients to health states in the model</u>

In a patient simulation state-transition model, for each cycle the probability of each transition probability is compared with a random number to determine if that particular transition occurs in that cycle. When the random number is less than the probability, the transition element is set to 1 thus indicating that the transition occurs. If the random number is higher than the probability then that specific transition does not occur.

The ERG found a mistake in the formulae used to allocate patients to the different health states in the model.

To model the 3 health states for which transition probabilities are dependent on time since transplantation (STP, AR and MCR) the formulae need to consider time elapsed since transplantation. However in Novartis' excel model, what is driving the allocation of patients into health states is time spent in a specific health state, not time since transplantation. This means that every time a patient changes heath state, time "restarts" in the model. Whilst this is correct for constant transition probabilities or transition probabilities dependant on time previously spent in a health state, it doesn't properly account for time since transplantation. The ERG provide an example to aid the illustration of the problem.

Table 39 illustrates how patients are allocated to the SPT and the AR heath states in the EVR+rTAC arm of the economic model. As previously explained, if the transition probability is higher than the random number (column 7), then the transition will occur. The transition probabilities have been made dependant on the number of cycles in the state (last column), therefore we can see that in the example provided, the probability of a patient experiencing an AR event is the same in month 6 and month 15 (column 5), because for both these occasions this simulated patient has been on the SPT state for 1 cycle. However, we know that after 1 year post-transplantation the probability of a patient experiencing an AR episode is supposed to be 2.45% and not 1.90% (Table 22).

The same problem happens for the MCR state, where patients are only allowed to move to the state after 1 year of being in the SPT state (and not 1 year after transplantation as the company claim).

The implication of this is quite a significant one:

Looking at Table 40 we can observe that the probability of having an AR at 6 and at 9 months is higher than the probability of experiencing an AR episode at month 3 for the EVR+rTAC arm of the model. So this means that the allocation of patients to the AR state in the intervention arm of the model is actually being underestimated (note that in the example below, the probability of a patient experiencing an AR episode at year 1 should be higher than 1.9%).

However, when looking at the transition probabilities used in the comparator arms of the model, there is a decrease in the probability of having an AR episode from the end of month 3 to the end of month 6. This means that there is a significant overestimation of AR episodes in the comparator arms of the model. Making the analogy with the example given for the EVR+rTAC of the model, by year 1 a patient in the AZA+TAC arm of the model should have a 1.55% probability of experiencing an AR episode. However, the way the model currently functions leads to the patient having a 13.8% transition probability of AR at year 1.

Whilst the direction of the impact of this mistake in the model outcomes is straightforward (i.e. it favours the EVR+rTAC arm of the model), the quantification is not easily foreseen. This structural error is likely to only impact second episodes of AR (note that the probability of first AR in Table 39 is correct, the problem is the probability of experiencing subsequent episodes) and the probability of AR is not very high in any case.

The fact that the probability of experiencing a AR episode increases from 3 months to 6 and 12 months in the EVR+rTAC of the arm is explored in Section 5.2.3.

The bias induced by retarding the transition to MCR is also not easily quantifiable, but given that the probability of patients going from SPT to MCR is the same across treatment arms, this should not have a great impact on the final ICER.

Finally it should be noted that this problem doesn't affect the renal sub-model. The transition probabilities in the renal model are only dependent to the immunosuppressive treatment received by the patient, and are not time dependent.

This problem raises the previously mentioned question of the appropriateness of using a patient simulation state-transition model. Not only is it unlikely that these problems would have arisen if a cohort model had been used, but also the computational burden of the Novartis model does not allow the ERG to easily make changes to the model structure and estimate the impact of correcting mistakes made by the company in their model.

Table 39. Example of allocation of patients to health states in EVR+rTAC arm of Novartis model

Cycle no.	Ends after	Stable PT state		Acute rejection state		Random #	Current state	Complete cycles in state	Number of cycles in the state
		Transition probability	In the state?	Transition probability	In the state?				
1	Month 3		1			65.3%	Stb PT	**0**	1
2	Month 6	98.10%	0	**1.90%**	0	63.1%	Stb PT	**1**	2
3	Month 9	96.30%	1	**3.70%**	1	97.2%	AR	**0**	1
4	YEAR 1	98.12%	1	**0.00%**	0	26.4%	Stb PT	**0**	1
5	Month 15	98.10%	1	**1.90%**	0	65.3%	Stb PT	**1**	2
6	Month 18	96.30%	1	**3.70%**	0	71.1%	Stb PT	**2**	3

Table 40. Probability of AR used in the economic model

Variable	Everolimus + reduced dose tacrolimus	Azathioprine + standard dose tacrolimus	MMF + standard dose tacrolimus
Probability of AR 3 months	1.90%	13.80%	11.30%
Probability of AR 6 months	3.70%	7.50%	4.10%
Probability of AR 12 months	2.45%	1.55%	0.6%

It is also worth mentioning that in a patient simulation state-transition model, each row of the transition matrix must still sum to one, just as it must for a cohort-level state-transition model. So

if more than one transition in a row is sampled to occur a rule must be set up to determine which of the possible transitions is acted on (Davis et al, 2014). Novartis have assumed that deaths have precedence over other events.

According to David et al,2014 this can create a bias towards the more extreme event which ensures that rarer but potentially more dangerous events are not ignored within the model but this means that the model is likely to favour treatments which prevent the more severe events over those that prevent the less severe events. The bias that this generates within the cost-effectiveness estimates can be minimised by reducing the cycle length, as this lowers the probability that two events are sampled to occur within a single cycle.

Superseded
See Erratum

Cycle length

The cycle length in the economic model is 3 months and a half-cycle correction was not applied.

The ERG are concerned that 3-month cycles could not capture all the relevant outcomes for the disease modelled as it seems that disease progression is faster in real life than in the model. The company submission sated that for example, if patients failed the first steroid treatment in the AR state, this would be repeated 3 days later with the same success probability. Therefore, the ERG have raised this as a question of concern during the clarification stage.

The company explained that *the 3-month cycle length was validated with clinical experts who advised that the timing of any acute rejection was likely to occur within the three month period, but that the majority of patients would move back to the stable state after treatment*.

Novartis added that clinical opinion was that AR and ASRR are mainly asymptomatic stages, *where no additional healthcare costs would be incurred or benefits be gained* compared with the stable state, thus they would not expect that shortening the cycle length would have an impact on the results of the model, *as the costs of the acute rejection state are the same as the SPT state (with the exception of the relatively small cost of high dose corticosteroids) and the utility is also the same.*

The clinical opinion provided by the ERG's clinical advisor is that 3-month cycles seem too long to capture all the relevant events and that monthly cycles would perhaps be more appropriate. Furthermore the ERG find Novartis' argument that costs and benefits associated with the AR and ASRR are similar to those in the SPT state quite surprising as this is not the case. Not only does the submission report very different resource use for these states but this is also the case for the economic model, where ASRR costs (£6,506) are ninety-fold the SPT ones (£72) and AR costs (£1,922) are twenty-seven-fold the SPT ones.

Finally, from a methodological point of view, a 3 month-cycle is a relatively long one, thus a half-cycle correction should have been applied. Even though the ERG wouldn't expect the impact on the model outcomes to be significant, no justification was given by the company as to why this wasn't done. Furthermore, as mentioned before, when running patient simulation models short cycles are also beneficial to avoid likely bias. Davis et al, 2014 suggests that the cycle length

should be shortened to a value where multiple transitions to clinically or economically significantly different states within one cycle are extremely unlikely.

Time horizon

The time horizon considered in the economic model was 320 cycles (80 years). Given that the average staring age of patients is 54 years (with a standard deviation of 10 years to allow for random sampling of this parameter), the time frame seems unnecessarily high, adding to the computational burden of the model. After 80 years in the model these patients will be, on average, 134 year old. After 40 years in the model (i.e when patients are on average 94 years), 100% of patients are virtually dead.

Superseded
See Erratum

Number of simulations

The model simulated 10 000 individual patients. The ERG didn't find any justification for the number of patients selected, or any mention to this parameter throughout the submission. Methods of justification can include a graphical representation of the costs, QALYs and the cost per QALY gained and determining at what number of patients the estimated standard error in the results appear acceptable (Davis et al, 2014).

If the results are found to vary significantly when selecting a different random number stream then the model should be checked to see whether there are any unintended correlations between samples that are supposed to vary independently.

The ERG are concerned with the number of simulations and the lack of stability in the patient simulation model. The ERG have run the Novartis simulation model for the base case as submitted by the company, **without any changes,** to test the model stability with regards to ICER results. After running the model two times, the ERG found a considerable variation in the ICERs reported (Table 41), especially in the case of the EVR+rTAC versus MMF+TAC ICERs, which varied between £110,797 and £120,651 (nearly a 9% change).

To also note is that the variation observed in the ICER for EVR+rTAC versus AZA+TAC is smaller than the one observed in the ICER for EVR+rTAC versus MMF+TAC arm. Nevertheless it is still a significant one.

The ERG tried to run the model with 15,000 simulations but not only did the company made this option not possible by default in the model (the cell input value for the number of simulations was set to allow a maximum of 10,000) but also, when the ERG changed this definition and ran the model with 15,000 simulations the model break near to the end of the simulation process.

Table 41. Simulations ran by the ERG – no changes to the company model

Cost-effectiveness results per patient	EVR+rTAC (1)	AZA+TAC (2)	MMF +TAC (3)	Incremental value (1-2)	Incremental value (1-3)
Total costs £	**140,750**	**105,529**	**102,746**	**£35,221**	**£38,004**
QALYs	**4.210**	**4.022**	**3.867**	**0.188**	**0.343**
ICER (base case reported in submission)				**£187,842**	**£110,797**
Total costs £	140,509	106,279	105,097	£34,230	£35,412
QALYs	4.205	4.019	3.903	0.816	0.302
ICER (base case on first simulation ran by the ERG)				**£184,476**	**£117,216**
Total costs £	138,400	103,700	102,930	£34,700	£35,470
QALYs	4.173	3.982	3.879	0.191	0.294
ICER (base case on second simulation ran by the ERG)				**£182,003**	**£120,651**

In conclusion, the ERG found that the reporting of the model structure and its assumptions lacked clarity. The approach undertaken raises the following concerns:

- The ERG are not convinced that a patient simulation state-transition model is required in this situation. The company haven't reported sufficient details or a solid justification for their approach. It seems that a traditional cohort state-transition model would be appropriate in this case.

- The ERG believe that more emphasis should have been placed on the renal component of the economic model. More specifically, interaction between the 2 models should have been considered as it is known that not only immunosuppressive therapy for liver post-transplantation has an impact on renal functioning but also, equally important, that renal functioning has an impact on graft survival.

- In terms of the model structure, the ERG question the clinical plausibility of the MCR state and consider the cycle length too large to reliably capture clinically important events.

- We have identified errors in the formulae allocating patients to different health states in the model which:
    - Are likely to have been avoided if cohort state-transition model had been used

- Could not be corrected by the ERG due to the computational burden of running the simulation model
- Are likely to bias the results in favour of EVR+rTAC in the hepatic model.

- The number of simulations in the model doesn't seem to be enough to obtain stable ICERs. This might be due to a number of factors. Nonetheless, this limits the validity of the results of the economic analysis. This is further discussed in Section 6.

5.2.3 Data inputs

Overall the ERG find that the presentation of data inputs throughout the submission lacked clarity and transparency. Equally, as the model used so many parameters taken from available literature, a more thorough justification for the choice of these sources would have been helpful to ascertain the quality of the evidence produced by the model.

- #### Patient group

Novartis model used baseline patient characteristics to simulate a range of different patients going through the model. The baseline patient characteristics used are presented in Table 42.

In order to obtain random samples of the baseline values for each patient, Novartis assumed a normal distribution for all the parameter, with the respective mean and standard deviation shown in the second and third column of Table 42.

H2304 population

All the modelled patient groups are reflective of the population on the H2304 study as the mean characteristic values were reported to have been taken from the trial. This means that MMF and AZA patients were assumed to have the same baseline characteristics as EVR patients with respect to the variables presented in Table 42.

Looking at Figure 10, the mean age and respective SD values seem to be in accordance with the H2304 population. The gender in the trial was composed of 74% males, while in the economic value the mean was 50%, with the upper bound of the CI being 60%.

The ERG note that the company has used body weight instead of BMI, nonetheless BMI (and not weight) was the baseline value reported in H2304 and all the comparator studies, and it was also identified by Novartis as being a potential treatment modifier.

Finally, the variation in eGFR in the H2304 is significantly lower than the one considered in the economic model. This is further explored in the next subsection.

Comparator studies' populations

The weighted average age across all comparator studies is 57 years, while the gender distribution is in favour of males (63%). The eGFR weighted average value in the 2 comparator studies which reported eGFR at baseline (McDiarmid et al, 1993 and Porayko et al, 1994) is 83.4.
Overall, for the baseline characteristics simulated in the economic model, it seems that, on average, the respective values were not dissimilar across intervention and comparator studies. The exception is BMI, for which the ERG can't make an assessment given the estimate used in the economic model was body weight, and not BMI.

Nonetheless, study populations varied significantly with respect to other characteristics. For example, HCV prevalence at baseline varied between 7% (in Boudjema et al, 2011) and 100% in 3 studies.

There was also significant variance across potential treatment modifiers across the all studies. Mean donor age varied from 33 to 52 years, while mean cold ischaemic time varied from 7 to 12 hours (with a very large proportion of studies not reporting these). Discussions with our clinical expert indicated that the 'Donor Risk Index' would have provided a better measure for organ quality overall. A large proportion (60%) of studies failed to report the baseline prevalence of diabetes and for the studies which reported this value the variance was quite significant going from 0% in the Boudjema et al, 2011 study to 42% (in H2304). Similarly, hypertension was reported only for 23% of the comparator studies and values ranged from 4% to 35% (H2304) of patients having the condition at baseline.

It is worth noting that:

- From all the potential treatment effect modifiers identified by the company during the NMA analysis (donor age, cold ischaemic time (organ quality), HCC, exclusion of concomitant steroids in the first 3 months, diabetes and hypertension) none have been included in the simulation model.

- A subgroup analysis undertaken in the H2304 trial (reported in Section 4.2) showed no statistically significant relationships between age, gender, race, eGFR, HCV status, MELD score and cause of ESLD and the composite clinical endpoint (tBPAR, graft loss or death). The same conclusion was obtained for age, gender, race, region, hepatitis C serology status, eGFR strata, MELD categories and ESLD and renal outcomes.

Interestingly, only age, gender, body weight and mean eGFR were included in the simulation model. To note is also the fact that age, gender and mean eGFR did not seem to be very different across study populations.

Without any kind of justification given by the company as to why these were the selected baseline characteristics to be simulated in the model (and as mentioned in section 5.2.1), the ERG conclude that perhaps these were not the most appropriate ones to be included, or that other baseline characteristics should have been included. This could have justified the

implementation of a patient simulation model, which in the present analysis is not entirely justifiable.

Table 42. Baseline population in the economic model – all treatment arms

Parameter	Mean values used in the model	Standard deviation (SD)	Low	High
Age	53.8	9.58	18.0	70.0
Gender	0.5	-	0.5	0.6
Body weight (Kg)	73.9	14.99	59.1	88.7
Mean eGFR	81.0	47	17.5	259.1

Figure 10. Baseline population in H2304

	EVR+reduced TAC N = 245	TAC elimination N = 231	TAC controls N = 243
Age (years)	53.6 ± 9.2	53.2 ± 10.8	54.5 ± 8.7
Male gender, n (%)	180 (73.5)	164 (71.0)	179 (73.7)
Race, n (%)			
Caucasian	211 (86.1)	196 (84.8)	195 (80.2)
Black	4 (1.6)	6 (2.6)	9 (3.7)
Asian	4 (1.6)	8 (3.5)	5 (2.1)
Other	21 (8.2)	17 (7.4)	28 (0.7)
Missing	5 (2.0)	4 (1.7)	6 (2.5)
Body mass index (kg/m^2)[1]	25.1 ± 4.2	25.3 ± 4.3	24.5 ± 4.2
HCV positive, n (%)	78 (31.8)	72 (31.2)	76 (31.3)
eGFR (MDRD4) (mL/min/1.73^2)[1]	80.8 ± 32.7	82.9 ± 37.2	78.9 ± 27.7
Diabetic n (%)[1]	95 (38.8)	83 (35.9)	101 (41.6)
Primary disease leading to liver transplantation, n (%)			
Alcoholic cirrhosis	71 (29.0)	49 (21.2)	51 (21.0)
Hepatitis C	62 (25.3)	56 (24.2)	57 (23.5)
Hepatocellular carcinoma	42 (17.1)	31 (13.4)	35 (14.4)
Hepatitis B	17 (6.9)	17 (7.4)	15 (6.2)
Sclerosing cholangitis	8 (3.3)	20 (8.7)	12 (4.9)
Primary biliary cirrhosis	8 (3.3)	11 (4.8)	8 (3.3)
Metabolic disease	5 (2.0)	4 (1.7)	6 (2.5)
Cryptogenic cirrhosis	7 (2.9)	11 (4.8)	18 (7.4)
Autoimmune hepatitis	4 (1.6)	7 (3.0)	6 (2.5)
Acute hepatic failure	2 (0.8)	2 (0.9)	3 (1.2)
Other	19 (7.8)	23 (10.0)	32 (13.2)
MELD score[2]	19.2 ± 9.0	19.6 ± 7.5	19.0 ± 7.6
Donor age (years)	48.8 ± 18.2	50.0 ± 18.2	48.7 ± 17.4
Cold ischemia time (h)	8.4 ± 4.4	7.5 ± 2.7	8.0 ± 5.2
Acute rejection prior to randomization, n (%)			
tBPAR	15 (6.1)	10 (4.3)	13 (5.3)
BPAR	20 (8.2)	16 (6.9)	20 (8.2)
Acute rejection[3]	21 (8.6)	20 (8.7)	24 (9.9)

[1]At randomization.
[2]MELD score based on laboratory values only.
[3]Clinically suspected acute rejection regardless of biopsy confirmation.
Continuous variables are shown as mean (SD).
BPAR = biopsy-proven acute rejection; tBPAR = treated biopsy-proven acute rejection; eGFR = estimated GFR; HCV = hepatitis C virus; MDRD4 = abbreviated modification of diet in renal disease; MELD = model for end-stage liver disease.

- ## Clinical effectiveness data

The main source for clinical effectiveness data was the H2304 trial for the EVR+rTAC arm of the model, complemented with data from the studies included for the comparator set. These data were used to undertake an NMA analysis, which produced the outputs used to inform the calculation of transition probabilities within the model. Additionally the model used several parameters taken from available literature.

In this subsection we focus on the model parameters used in the hepatic and the renal models and describe how transition probabilities between health states were estimated within the different arms of the main economic model.

The ERG have several concerns with the clinical parameters used in the economic model.

Superseded
See Erratum

H23014 study – TAC trough levels

The main source for the effectiveness of EVR+rTAC is H2304 trial. Novartis acknowledge that one of the limitations of the trial is the fact that the target ranges of TAC are higher than the ones observed in standard UK clinical practice (where through levels <5 ng/mL are considered as low-dose TAC). This is attributed to the trial being set in the US and changes in clinical practice since 2007.

As previously mentioned, in the EVR+rTAC arm of the trial the planned dose of TAC was to achieve the 3 – 5 ng/mL by 3 weeks after randomization and keep on these for the remainder of the study. In the standard TAC or the trial, the trough levels were targeted to be maintained at 8-12 ng/mL until month 4, and at month 4 TAC whole blood trough levels would be decreased to a target trough level of 6-10 ng/mL for the remainder of the study. Finally, in the TAC elimination arm (i.e. the EVR arm), after EVR blood trough levels were confirmed to be in the target range (3-8 ng/mL), TAC tapering was started, achieving a target TAC trough level of 3-5 ng/mL by 3 weeks after randomization. TAC elimination began at month 4 post-transplant and TAC was completely eliminated by the end of month 4.

In Figure 11 below the ERG present the TAC trough levels reported in the CSR for the 12-month analysis.

It can be observed that in general the average trough levels of TAC were higher than the ones initially planned for all arms of the trial. In fact the “reduced” TAC arm shows trough levels above 5 ng/mL throughout the 12 months. Clinical opinion sought by the ERG indicated that the trough levels observed in the reduced TAC arm are comparable to what would be considered a standard TAC regimen in the UK. This means that the standard TAC arm in the trial is also not reflective of UK clinical practice, presenting extremely high trough levels. Our clinical advisor explained that a standard TAC regimen in the UK is 6 – 8 ng/mL until month 1, just above 6 ng/mL until month 4 and then between 5 and 6 ng/mL until the end of first year.

It is also noticeable that in the TAC elimination arm, the average TAC trough levels after month 4 ranged from [redacted] and that while some patients seemed to be at 0 trough levels, this was not the case for all patients. It is therefore possible to hypothesise that the TAC trough levels in the TAC elimination arm are closer to what would be considered a reduced level of TAC in the UK than any other trial arms in H2304 (at least from month 4 to month 9). This leads to the question: “would the TAC elimination arm, and its outcomes, be more appropriate to capture the advantage of EVR (i.e. reduced levels of TAC) and its renal sparing effect within the UK clinical reality?”.

Figure 11. TAC trough levels by visit – local lab data, safety population – 12 months

[redacted]

The fact that the EVR+rTAC of the arm isn’t reflective of clinical reality in the UK in terms of TAC levels is worrying as it questions the validity and the applicability of the economic analysis within the NHS context.

Core hepatic model

NMA estimates

In the hepatic model, treatment effectiveness only impacted the patients' transition from the SPT to the AR state thus the type of immunosuppressive treatment that the patient receives (i.e. EVR+rTAC or any of the comparator regimens) is only relevant to determine the likelihood of the patient experiencing an AR episode.

All other transition probabilities were assumed constant and independent of the immunosuppressive treatment that patients receive.

In their submission the company reports that the probability of progressing from the stable post-transplant to the acute rejection state was determined by the probability of BPAR at 3, 6 and 12 months for each comparator, calculated in the NMA. However, no further details were given as to how the NMA BPAR data (originally presented as odd ratios) were used to derive time dependent transition probabilities presented in Table 54 in the submission.

Upon the ERG's request for further detail on the aforementioned calculations, Novartis provided the tables shown in Figure 9 (presented in Section 5.1) and reported that these are the *absolute results for the outcomes presented in the NMA.*

Novartis response to the ERG request for clarification is not really satisfactory as no additional information was provided by the company. Novartis simply presented the same results for which the ERG requested clarification and mentioned these are the absolute results from the NMA (our concern is explained in detail in the NMA methodology section).

Furthermore it should also be noted that the values presented are not the final model inputs, as the clinical effectiveness data were used to ultimately model the probability of an AR episode (reported by the ERG in Table 40) and not the probability of being tBPAR free. This should have been mentioned in the company submission. The ERG analysed these calculations in the economic model and present these below:

Using the probability of tBPAR free for EVR+rTAC at 3 months of 98.1% obtained from the NMA, the company estimated: (1- 98.1%) = 1.90% as the probability of AR at 3 months.

Using the probability of tBPAR free for EVR+rTAC at 6 months of 94.4% obtained from the NMA, the company estimated: (98.1% - 94.4%) = 3.70% as the probability of AR at 6 months.

Using the probability of tBPAR free for EVR+rTAC at 12 months of 89.5% obtained from the NMA, the company estimated: (94.4% - 89.5%) / 2 = 2.5% as the probability of AR at 9 months and 2.5% as the probability of AR at 12 months.

As the economic model uses 3-month cycles, and the NMA results were provided for the 3, 6 and 12 months periods, the company had to undertake additional calculations to estimate the 9-month probability of AR. However, the calculations undertaken are essentially wrong from a methodological point of view as probabilities cannot be divided by a time period t, in order to get the cycle probability for a t-length cycle (Briggs et al, 2006). Probabilities need to be converted to rates first and then converted to probabilities again, assuming the event occurs at a constant rate. As the probability of AR is small (5%), the impact of correcting the company calculations is not a great one:

Probability of being tBPAR free at 6 months is: 94.4%
Probability of being tBPAR free at 12 months is: 89.5%

Rate of AR: - [ln (1- (94.4% - 89.5%))] / 6 months = 0.0084
3-month probability assuming a constant rate of AR: 1 – exp (-0.0084 * 3) = 2.48%

Even though this only has a very small impact on the final estimate (and extremely unlikely to have an impact on the model outcomes), if the probability of AR was higher, there would be a considerable difference in the final values. The ERG found this mistake in other calculations. We explore these in this section.

Similarly to EVR+rTAC, the impact of correcting these calculations in the AZA+TAC and the MMF+TAC arms of the model is unlikely to have an impact on the model outcomes, as the change in values is extremely small.

Looking at Table 40, it can noted that there is a decrease in the probability of having an AR episode from the end of month 3 to the end of month 6. This means that there is a significant overestimation of AR episodes in the comparator arms of the model. Making the analogy with the example given for the EVR+rTAC of the model, by year 1 a patient in the AZA+TAC arm of the model should have a 1.55% probability of experiencing an AR episode. However, the way the model currently functions leads to the patient having a 13.8% transition probability of AR at year 1.

Finally, the fact that the probability of experiencing an AR episode in the model in the EVR+rTAC of the arm is higher at 6 and 12 months than the same probability at 3 months lacks face validity. As mentioned throughout the submission the probability of occurrence of AR decreases as time post-transplant goes by. Of note is that the probability of AR in the MMF+TAC and in the AZA+TAC arms of the model shows a decreasing trend.

<u>Literature estimates</u>

Table 22 in Section 5.1 presents the transition probabilities taken from the evidence base. The ERG found several issues related with the estimates used. These are now explored, together with the appropriateness of the sources used. More specifically the following transition probabilities are discussed:

- SPT to MCR (stable post-transplant to mild chronic rejection)
- AR to SPT (acute rejection to stable post-transplant)
- ASRR to SPT (acute steroid-resistant rejection to stable post-transplant)
- ASRR to SCR (acute steroid-resistant rejection to severe chronic rejection)
- MCR to MCR (staying in the mild chronic rejection state)
- SCR to SPT (severe chronic rejection to stable post-transplant)
- SCR to death (severe chronic rejection to death)

SPT to MCR (stable post-transplant to mild chronic rejection)

The company reports to assume that this transition can only occur 1 year after transplant, however the ERG found a mistake in the model with regards to the allocation of patients to the health states (discussed in Section 5.2.2 of our report) which means that patients can only move to the MCR state 1 year after being in SPT state (and not 1 year after transplantation).

The company reported this value to be 4% as per clinical opinion. However the value used in the economic model is 1%. The company make no mention to the calculation involved in obtaining this value but the ERG assume this is 4%/4 months = 1% every 3 months. The same issue raised above for dividing transition probabilities by time intervals applies here, even though the final result is virtually the same when using the correct method.

Surprisingly, with 1% of patients moving to this heath state every month, no apparent clinical plausibility to justify its existence and no determent in QoL associated with it, according to the company sensitivity analysis **when the MCR state is eliminated from the analysis, the final ICER comparing EVR+rTAC with AZA+TAC is £176,410, while the ICER for EVR+rTAC compared with MMF+TAC is £233,331**. (Of note is that base case ICERs are £187,842 for EVR+rTAC vs AZA+TAC and £110,797 for EVR+rTAC vs MMF+TAC). This is further explored in Section 6.

AR to SPT (acute rejection to stable post-transplant)

The company reports using the 86.3% probability of successful first line steroid therapy and source the value back to Aydogan et al, 2010. However, in the economic model the probability used to account for patients going from AR to SPT is 98.1%. Novartis arrive to this estimate by undertaking the following calculation:

1 – [(1 – 86.3%) ^ (number of steroid treatment courses)]. As they assume number of steroid treatment courses = 2, then:

1 – [(1 – 86.3%) ^ (2)] = 98.1%.

From a conceptual point of view, the formula can be translated as:

1 – [(probability of failing steroid treatment) * (probability of failing steroid treatment)] which is equivalent to the probability of success of 1st line steroid treatment + probability of 1st line steroid treatment failing and second line steroid treatment being successful.

However, two things should be consider:

- The 86.3% of successful steroid treatment in Aydogan et al, 2010 is for a 1 course (3 days) treatment. Nonetheless Novartis assume the same probability of successful resolution of AR for 1st and subsequent episodes.
- The costs associated with the effectiveness of steroid treatment in the model were accounted for the average of 1.4 courses per patient which is the equivalent to the number of therapies considering the probability of patients failing (or succeeding) 1st line and then moving on to a 2nd line of steroid treatment. So even though Novartis assume that patients can get up to 3 consecutive therapies (as per their Table 50) the base case analysis only considers 2 possible lines of treatments.

The fact that the probability of successful steroid treatment for AR (up to two treatments) is nearly 100% in the simulation model is the reason why virtually no patients experience episodes of ASRR in any of the treatment arms.

Superseded See Erratum

Clinical opinion sought by the ERG informed us that the overall probability of successfully treating episodes of AR with steroids is around 90%.

ASRR to SPT (acute steroid-resistant rejection to stable post-transplant)

The company assume an 83.3% probability of successful treatment for ASRR which was taken from Aydogan et al, 2010. In the study, patients received 5mg/kg/day of anti-thymocyte globulin treatment for an average of 11 days (5 - 14 days range). However, in their economic model Novartis use 1.5 mg/kg/day of anti-lymphocyte therapy. Clinical opinion sought by the ERG agreed that the effectiveness of anti-lymphocyte therapy is expected to be similar to the effectiveness of anti-thymocyte therapy.

ASRR to SCR (acute steroid-resistant rejection to severe chronic rejection)

The company assume a 16.7% probability of unsuccessful treatment for ASRR which was taken from Aydogan et al, 2010. Patients then progress to SCR state, which is consistent with the study. The treatment regimen associated with the successful (or unsuccessful) resolution of ASRR in the study is different from the treatment assumed in the economic model as mentioned in the previous point.

SCR to SPT (severe chronic rejection to stable post-transplant via re-transplant)

In the economic model patients are assumed to stay in the SCR state for 2 cycles (6 months) and are re-transplanted after this period. Novartis state that this assumption is based on the Organ Donation liver statistics (2014), which report the median time spent on the liver transplant

waiting list. This estimate is 145 days (134 – 156 days with a 95%CI) and not 147 as per the company submission.

145 days is the equivalent to nearly 5 months. However the company choose a 6 months waiting period. Again, this might reflect the incorrect choice of cycle length, where shorter cycles would allow to appropriately capture this waiting period.

As is the case for other parameters in the model, the estimates used in the SCR state of the model (and the issues related to these) ultimately do not have an impact on the final base case results since there were virtually no episodes of SCR or no re-transplants in the model.

Data parameters used to model mortality in the hepatic model are explored in the mortality subsection.

Renal model

<u>NMA estimates</u>

In the renal sub-model, treatment effectiveness has an impact on patient's renal functioning for the first year of the model. Renal dysfunction was measured through the decrease in patients' baseline level of eGFR.

The transition probabilities in the renal sub-model define the progression of CKD for each patient. While in the hepatic core model all patients are assumed to start in the SPT state, in the renal model the patient starting level of CKD is randomly generated through a simulation and the decrease in renal function (over the first year) is dependent on the treatment regimen. We now provide more insight on how the baseline CKD level is simulated:

The patient baseline eGFR level is randomly obtained. This is also the case for other patient baseline characteristics. In order to obtain random samples of the eGFR baseline level for each patient, Novartis assumed a normal distribution for this parameter, with a mean of 81 mL/min per 1.73 m2 and a SD of 47 mL/min per 1.73 m^2. The company explains in the submission that all the baseline characteristics were taken from the H2304 population.

Looking at Table 42 previously presented, it can be noted that the eGFR baseline level in the EVR+rTAC arm of H2304 was 80.8 mL/min per 1.73 m^2 with a SD of 32.7 mL/min per 1.73 m^2. Hence, the SD of 47 mL/min per 1.73 m^2 chosen by the company to use in the model seems out of line with the values reported in the CSR. Even though the simulation in the economic model produced a baseline eGFR average level of 82 mL/min per 1.73 m^2, the SD used is not justifiable.

After simulating the patient baseline eGFR level, the expected 12-month decrease in eGFR for the specific treatment arm (reported in Table 23) is applied to the baseline value. Depending on the final eGFR value, the patient is allocated into a CKD category (from no CKD to CKD stage

5). This allows the estimation of the impact of the specific immunosuppressive therapy (i.e. EVR+rTAC or the comparators) on renal functioning by the end of the first year post-transplant.

The way the simulation was run in the model produces clinical implausible results. Due to the random generation process used to estimate the baseline eGFR levels and the use of the normal distribution, 4.2% of patients will start the model with negative levels of eGFR, which doesn't seem realistic. This means they are immediately allocated to the CKD stage 5 category, where these patients with a negative eGFR level can stay for long periods of time (like for example over 3 years) until they get transplanted and back to CKD levels 1- 2. Clearly this is isn't a plausible representation of clinical reality.

Novartis didn't provide details as to how the NMA eGFR data (originally presented as the difference in eGFR change from baseline) were used to derive the eGFR decrease at 12 months for the different treatment regimens used in the economic model, presented in Table 23.

Upon the ERG's request for further detail on how the company used the NMA data to obtain the estimates in Table 18, Novartis provided the tables shown in Table 24 and reported that these are the *absolute results for the outcomes presented in the NMA.*

Novartis response to the ERG request for clarification is not really satisfactory as no additional information was provided by the company. Novartis simply presented the same results for which the ERG requested clarification and mentioned these are the absolute results from the NMA (our concern is explained in detail in the NMA methodology section).

Novartis also decided to model the decrease in eGFR at 12 months in the comparator arms of the economic model by using the decrease in eGFR at 12 months reported for standard TAC in the NMA (Table 24). Novartis justify this by arguing that *as ciclosporin has limited use in the UK (approximately 5% of current market - Novartis data on file 2014), and the comparators of interest are used in combination with standard dose tacrolimus, only the absolute value for standard dose tacrolimus was used in the economic model*.

The ERG have serious concerns not only with the validity of the estimates presented in Table 24 but also with the decision to use the standard TAC estimate to model the decline in renal function in the MMF and AZA arms of the economic model. This is now discussed.

Table 43 presents the trough levels of TAC in the different NMA studies. Analysis of these data raised the following issues:

- The ERG are deeply concerned with the fact that the MMF arms of the trials were considered by the company to include reduced doses of TAC. The trough levels of TAC in the "reduced" arms of Neuberger et al, 2009 and Boudjema, et al 2011 studies are similar to the trough levels of standard TAC in H2304. The average trough levels in the standard TAC arms in both studies are generally higher than the trough levels in the standard TAC arm of H2304. This means that the "reduced" TAC arms in these studies

are in fact standard TAC arms. And even if these are considered standard TAC levels, they are still higher than the ones verified in UK clinical practice.

- The ERG are equally concerned with the fact that the company decided to considered the levels of TAC in the Porayko et al., 1994 study as "standard" TAC. The TAC trough levels were maintained below 1.0 ng/mL for the entire year (beginning 4 weeks after randomization), which actually **represents the lower trough levels of TAC amongst all the of NMA studies (including H2304).** So in reality the study is comparing AZA+ ciclosporin with reduced TAC. The same is applicable to the McDiarmid et al, 1993 study, where TAC trough levels were maintained below 2 ng/mL.

Overall, the ERG question the validity of the NMA results for the renal outcomes. The allocation of different studies' treatment arms to the reduced and standard TAC categories is inconsistent and misleading.

Superseded
See Erratum

As the standard TAC connector across the NMA studies is so heterogeneous (note for example that in the Boudjema et al, 2009 study the trough level of TAC in the standard TAC arm was close to 12 ng/mL for several months during the first year) then the NMA results are likely to lack robustness.

Regardless of the ERG concerns with the validity of these estimates, looking at Table 24, it can be argued that the estimate for the decrease in renal function at 12 months for the MMF+ "reduced" TAC arm (28.2) should be used in the MMF arm of the economic model instead of the standard TAC estimate (31.6). This is because the MMF+ "reduced" TAC is actually closer to a MMF+ standard TAC arm if we take the H2304 study (and UK clinical practice) as reference. The same value could be used for the AZA arm of the model as the renal dysfunction is determined by the levels of TAC (and is not dependent on the concomitant drug).

Table 43. TAC trough levels in the NMA studies.

H2304	Standard TAC	The standard TAC arms presents average trough levels of 10 ng/mL at randomization, around 9 ng/mL until month 9 and then close to 8 by the end of the 1st year.
	EVR + reduced TAC	The reduced TAC arm presents average trough levels of 11 ng/mL at randomization and then around 6 ng/mL from month 4 until end of the 1st year.
(Neuberger *et al.*, 2009[65])	Standard TAC	The standard TAC arm presents average trough levels of 11 ng/mL until month 3 and then decreased to 9 by the end of the 1st year.
	MMF + reduced TAC	The reduced TAC arm presents average trough levels of 9 ng/mL until month 3 and then decreased to 8 by the end of the 1st year.

(Boudjema *et al.*, 2011[83])	Standard TAC	The standard TAC arm presents average trough levels above 10 ng/mL for the whole of the 1st year.
	MMF + reduced TAC	The reduced TAC arm presents average trough levels between 10 and 9 ng/mL for the 1st year.
(Porayko *et al.*, 1994[88])	Standard TAC	TAC trough levels were maintained below 1.0 ng/mL for the entire year.
	AZA+ciclosporin	
(McDiarmid *et al.*, 1993[93])	Standard TAC	TAC trough levels were maintained below 2 ng/mL.
	AZA+ciclosporin	

Literature estimates

Novartis assumed that at the end of year 1, patients' renal function follows a *natural progressive decline.* Table 20 in Section 5.1 presents the transition probabilities taken from the evidence base. These are now explored, together with the appropriateness of the sources used.

The main source used to model the natural progression of kidney disease was La Mattina et al, 2011. The company reported the values used and these are presented in Table 44.

However, when analysing the La Mattina et al, 2011 paper, the ERG concluded that the values reported by the company were the baseline values in the study. La Mattina et al, 2011 looked at the modifiable risk factors which may help prevent CKD progression in liver transplant recipients. The population in the analysis was defined as patients who undertook liver transplantation 1 year ago. The baseline CKD state of the study population is presented in Table 45.

Novartis approach is therefore questionable since:

- The company uses baseline proportions as annual progression rates. From a conceptual point of view the 2 are completely different. For example, if at time of transplant there were 52 patients at the CKD stage 1, and by the end of 1 year there were still 52 patients in the same state (which corresponds to 7% of the total population), then the annual progression rate would be 0. The values used by Novartis are not progression rates and cannot be used to estimate CKD progression.
- The company uses values measured at the end of year 1 post-transplant, but in fact the NMA analysis provided values for year 1 of the model, so Novartis need estimates of progression by the end of year 2 or longer periods.
- Novartis decided to aggregate CKD stage 1 and 2 in their analysis. However, La Mattina et al, 2011 differentiates the two. Instead, the company decided to use the estimate for CKD stage 1 in the original paper to model the no CKD state. Given that all the data were available to estimate the distribution across CKD stages separately, the ERG do not understand why the company decided to aggregate these stages.

In conclusion, Novartis should have used the values reported in Table 45 for the progression of renal disease at the end of year 3 by converting these into annual probabilities (using the method previously described by the ERG).

Table 44. CKD progression rates used by Novartis

State	Annual progression rate to next state (Lamattina *et al.*, 2011[116])
No CKD	7.0%
CKD stages 1/2	34.0%
CKD stage 3	56.0%
CKD stage 4	3.0%
CKD stage 5	-

Table 45. CKD baseline values and progression probabilities used in La Mattina.

State	CKD stage at 1 year (*n*, %) - baseline	Incidence of stage progression at year 3 based on CKD stage at 1 year	Incidence of stage progression at year 10 based on CKD stage at 1 year
CKD stage 1	52 (7%)	66%	89%
CKD stage 2	246 (34%)	49%	85%
CKD stage 3	406 (56%)	12%	33%
CKD stage 4	19 (3%)	6%	21%
CKD stage 5	6 (1%)	n/a	n/a

Source: Adapted from La Mattina et al, 2011

The values presented in Table 44 were used to estimate the 3-month cycle probability of disease progression after year 1 in the model (Table 25). The company reports that the *annual probability of progression was converted to rates and adjusted for cycle length, then re-converted to probability.* This is not true. In fact, this should have been the methodology taken (as previously explained by the ERG), but in reality the company simply divided the annual values by 4 to obtain cycle estimates.

For example, taking the 34% value used to model progression from CKD stage ½, Novartis derive the cycle probability of 8.5% (34%/4) shown in Table 25. However following the correct methodological approach yields a value of 9.9%. The difference in estimates is considerably higher than in previous cases, as the annual probability value is considerably higher (i.e. 34%).

As before, data parameters used to model mortality in the renal sub-model are explored in the mortality subsection

In conclusion, the ERG have several concerns related with the clinical effectiveness data used in the model:

- The ERG found mistakes in the calculations of effectiveness data in the hepatic model. Nonetheless, these are unlikely to impact the final model results as the corrected values are quite similar to the wrong ones in terms of absolute value and also some of the mistakes identified were in heath states that are virtually not populated in the model.

- The way the simulation was run in the model produces clinical implausible results. Due to the random generation process used to estimate the baseline eGFR levels, some patients start the model with negative levels of eGFR, which is not realistic. This means they are immediately allocated to the CKD stage 5 category, where these patients with a negative eGFR level can stay for long periods of time (like for example over 3 years) until they get transplanted and back to CKD levels 1- 2. Clearly this is not a plausible representation of clinical reality.

- Novartis used baseline proportions reported in the La Mattina et al, 2011 study as annual progression rates. The values used by Novartis are not progression rates and cannot be used to estimate CKD progression. The company uses values measured at the end of year 1 post-transplant, but in fact the NMA analysis provided values for year 1 of the model, so Novartis need estimates of progression by the end of year 2 or longer periods.

Overall, the ERG question the validity of the NMA results for the renal outcomes. The allocation of treatment arms from different studies to the reduced and standard TAC categories is inconsistent and misleading:

- The ERG are deeply concerned with the fact that the MMF arms of the trials were considered by the company to include reduced doses of TAC. The trough levels of TAC in the “reduced” arms of Neuberger et al, 2009 and Boudjema, et al 2011 studies are similar to the trough levels of standard TAC in H2304. The average trough levels in the standard TAC arms in both studies are generally higher than the trough levels in the standard TAC arm of H2304. This means that the “reduced” TAC arms in these studies are in fact standard TAC arms. And even if these are considered standard TAC levels, they are still higher than the ones verified in UK clinical practice.

- The ERG are equally concerned with the fact that the company decided to considered the levels of TAC in the Porayko et al., 1994 study as “standard” TAC. The TAC trough levels were maintained below 1.0 ng/mL for the entire year (beginning 4 weeks after randomization), which actually **represents the lower trough levels of TAC amongst all the of NMA studies (including H2304).** So in reality the study is comparing AZA+ ciclosporin with reduced TAC. The same is applicable to the McDiarmid et al, 1993 study, where TAC trough levels were maintained below 2 ng/mL.

As the standard TAC connector across the NMA studies is so heterogeneous (note for example that in the Boudjema et al, 2009 study the trough level of TAC in the standard TAC arm was close to 12 ng/mL for several months during the first year) then the NMA results are likely to lack robustness.

- ## Mortality data

Mortality data used in the model

Core hepatic model

In addition to the background mortality rate (age and gender dependent), Novartis assumed a positive mortality risk for patients waiting for a liver transplantation.

Organ Donation UK, 2014 estimates that 12% of patients die on the liver waiting list every year.

Novartis has divided this value by 4 to obtain the 3 month cycle probability of death, which lead to the 3% mortality estimate. Again, the ERG note that this is not the right approach to transform a 12-month probability estimate into a 3-month one. The correct methodology yields a final value of 3.15%, so similarly to previous estimates, there is not much difference in the final parameters obtained since the estimate is low (i.e. 12%)

Furthermore, for the period of re-transplantation the company considered an additional mortality increment of 0.7%, taken form Aylin et al, 2013. However, the 0.7% value reported in Aylin et al, 2013 is the 30-day mortality rate for inpatient admissions for elective surgery (which didn't include liver transplantation in the analysis). So not only this is a monthly rate, which by definition should occur every month in a 3 month cycle (even though Novartis consider this to be a 3-month probability) but the analysis didn't include liver (or any other organ) transplantation procedures.

As is the case for other parameters in the model, the estimates used in the SCR state of the model (and the issues related to these) ultimately do not have an impact on the final base case results since there were virtually no episodes of SCR or no re-transplants in the model.

Renal sub-model

In addition to the background mortality rate (age and gender dependent), Novartis assumed a mortality risk depending on the patient CKD stage.

.

The source used to model the incremental mortality associated with CKD was Keith et al, 2004. The company reported the values used and these are presented in Table 46.

The ERG analysed the Keith et al, 2004 study. The purpose of the 5-year study was to understand the natural history of CKD with regards to progression to renal replacement therapy (transplant or dialysis) and death in a representative patient population in the US. Therefore, the

study population did not consist of liver transplant patients. The population in the analysis presented co-morbidities such as hypertension, diabetes mellitus, coronary artery disease, congestive heart failure, hyperlipidaemia, and renal anaemia.

The values presented in Table 47 were used to estimate the 3-month cycle mortality risk for CKD patients (Table 46). Novartis used the original 5-year probabilities, divided the values by 5 and then again by 4 to obtain 3-month estimates. Again, this is methodologically wrong. The ERG estimates the values using the correct approach and present these in the second row of Table 47.

Similarly to the clinical effectiveness data, the company aggregated CKD stage 1 and 2 in their analysis. However, Keith et al, 2004 differentiates between them. Instead, the company decided to average the mortality estimate for CKD 1 and CKD 2. Given that all the data were available to obtain estimates for the CKD stages separately, the ERG do not understand why the company decided to aggregate these stages.

Furthermore, as the study did not report mortality rates for CKD stage 5, Novartis assumed these to be twofold the CKD stage 4 rates. This should have been verified by clinical opinion and should not have been referenced as Keith et al, 2004 as it was an assumption adopted by the authors of the Novartis model.

Table 46. CKD increased risk of mortality (over 3 months) used by Novartis

State	Increased risk of mortality in state (Keith *et al.*, 2004[117])
No CKD	+ 0.0%
CKD stages 1/2	+ 0.74%
CKD stage 3	+ 1.22%
CKD stage 4	+ 2.29%
CKD stage 5	+ 4.57%

Source: Adapted from Novartis' submission Table 48

Table 47. CKD mortality over 5 years reported in Keith et al, 2014.

End Point	GFR, 60-89; No Proteinuria (n = 14 202)	Stage 2 GFR, 60-89; Proteinuria (n = 1741)	Stage 3 GFR, 30-59 (n = 11 278)	Stage 4 GFR, 15-29 (n = 777)
5 year mortality	10.2%	19.5%	24.3%	45.7%
ERG calculation – 3 month mortality	0.54%	1.08%	1.38%	3.01%

Source: Adapted from Keith et al, 2014

Similarly to the hepatic model, for the period of renal re-transplantation the company considered an additional mortality increment of 0.7%, taken form Aylin et al, 2013.

The ERG identified some problems with the use of this parameter (described in the core hepatic subsection above). However, unlike the hepatic model, these have an impact on the renal sub-model as renal transplants events do occur in the model.

The surgical mortality is underestimated in the economic analysis since the original estimate is a monthly rate, which by definition should occur every month in a 3 month cycle but Novartis considered this to be a 3-month probability. However since the value is considerably small (0.7%) the impact of this on the final outcomes is likely to be insignificant.

Estimated mortality in the model

Markov models produce outputs which allow the estimation of the proportion of patients alive, or death, as time goes by (Markov traces).

In patient simulations models, because 1 patients is put through the model each time (instead of a cohort of people), the outcomes are the average time spent in a specific health state across all patients. This state can be death thus allowing the estimation of the average time patients spent in the death state (and by default in the “alive” state).

Novartis state in their submission that given their analysis is based on a patient simulation model, a Markov trace for each state is not possible. However, this is not true. By counting the number of events which took place before certain points in time, one can derive an approximation of a traditional Markov trace.

The ERG have done the analysis and present the results here in terms of mortality across the different arms of the economic model.

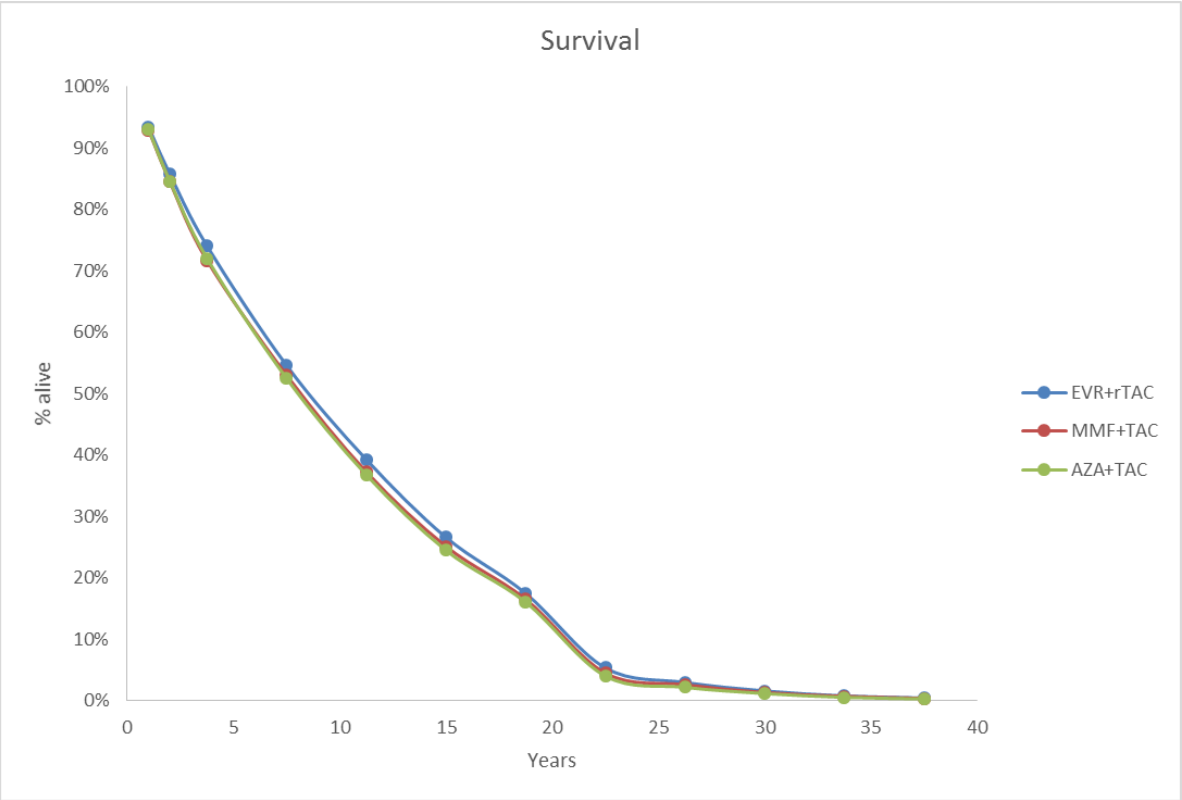
Figure 12 shows the proportion of patients alive in the 3 arms of the economic model as time goes by. It can be noticed that the difference in survival across the treatment arms is minimal, with survival in the EVR+rTAC arm only slightly above the other 2 treatment arms. So it can be concluded that incremental mortality is not one of the key drivers of the economic model.

As reported in the submission (and according to Barber et al, 2007), the median survival for patients between 24 and 54 years of age is expected to be 25.3 years (CI 20.5–31.2) whilst for 55 to 64 year old patients, the reported median survival is 19.5 years (CI 16.1–23.6).

To note is that in the economic model median survival is achieved around year 10 which indicates a considerable underestimation of survival in the model.

To also consider is that overall survival in the EVR+rTAC arm of H2304 in 95% by 2 years, while in the economic model only around 85% of patients are alive at the same point in time. Again, this indicates that survival is underestimated in the economic model.

Figure 12. Proportion of patients alive in the economic model



- Health related quality of life

Hepatic model

The ERG found a mistake in the reporting of the SCR utility value. In the submission, a utility value of 0.43 is reported for this health state however, the economic model uses a utility value of 0.53.

The SPT state yields a utility value of 0.58. Novartis state that patients have to undertake immunosuppressive therapy and that there is a loss in QoL associated with this kind of therapy, without further explanation. However, Novartis also consider, separately, the disutility associated with the AE related to the treatment regimens. Therefore if the administration and monitoring of EVR patients does not impact on patient's QoL, assuming a QoL loss beyond EVR-related AEs could be double-counting the QoL loss attributable to the immunosuppressive therapy.

The SPT state utility value was obtained by calculating a weighted average of the utility values reported in Ratcliffe et al, 2002 (Table 27) for different time points by the respective sample sizes. However, the values obtained from the QoL study reported a p-value of 0.55, which demonstrates a non-statistically significant difference in QoL values over time. So Novartis are using a range of values which are not statistically significantly different from each other over the period considered.

Novartis' clinical experts advised that the AR, ASRR and MCR state are asymptomatic. Therefore the company assumed that the QoL experienced in these health states is the same as the QoL experienced in the SPT state. However Novartis report that patients in these states might require hospitalization. Clinical opinion sought by the ERG agreed that these patients will require hospitalization and that this will likely lead to a reduction in their QoL. Therefore, the utilities obtained for the AR, ASRR an MCR are likely to be overestimated.

Finally, patients in the SPT (via re-transplant) state were also assumed to experience the same QoL as patients in the initial SPT state. As there were virtually no liver re-transplantations in the model, the impact of this assumption, despite its clinical plausibility (or lack of it) is limited.

Renal sub-model

It was assumed that the no CKD state yields a utility value of 0.83. It is stated that patients in this state suffer from no detriment in their QoL compared with stable hepatic patients. This value was based on an assumption, however this doesn't seem to have been validated by clinical expert opinion. Regardless, it is not clear why the value chosen was 0.83 since the stable hepatic patients experience a utility value of 0.58 in the model.

Novartis state that for CKD stage 1/2 patients observe a mildly reduced kidney function and that symptoms may include elevated blood pressure. The utility value associated with this state is 0.64.

Clinical opinion sought by the ERG indicated that the 0.83 utility value would be better suited to represent the CKD stage 1 and considered that the utility value of 0.64 is too low for the CKD

stages 1/2. It was added that the transition from stage 1 to stage 2 is unlikely to be significant in terms of change in patient's symptomology.

Utility vales used for CKD stage 3, 4 and 5 are reasonable according to the input of our clinical advisor.

Finally, patients in the no CKD state (via re-transplant) yield a utility value of 0.81. Novartis state that these patients experience a similar HRQoL to the SPT state in the main (hepatic rejection) model. However the value used is not similar to the one used for the hepatic model (0.58) hence the ERG assume this is a reporting mistake.

Relationship between the core hepatic model and the renal sub-model

Novartis' clinical experts advised that the different states in the core hepatic model and renal sub-model could be considered independent in terms of impact in QoL. The ERG find this statement quite surprising given the company previous claims that declining renal function can have a major impact on graft and patient survival and that in transplant patients, chronic kidney disease is associated with an increased predisposition to cardiovascular events, an increased risk of hospitalisation, hepatic allograft dysfunction and an increase in mortality when compared with patients with preserved renal function.

Therefore, and in line with the ERG previous comments, an interaction between the QoL in both models would be expected and clinically relevant.

The company excluded from the analysis health effects associated with hyperacute rejection, HCV and HCC as it was considered that the choice of immunosuppressive regimen has no impact on these outcomes (thus there wouldn't be differences between treatment arms). Nonetheless, the treatment regimen will impact on patient's survival. Therefore some patients will live longer than others (as a result of being in a specific treatment arm) and can experience HCV or HCC (and respective costs and QALYs) for different periods of time across treatment arms.

- ### Resources and costs

The ERG are generally satisfied with the sources used to obtain unit costs for the hepatic model.

Estimates of resource use were either based on assumptions or obtained from the University Hospitals Birmingham Foundation Trust hence the ERG undertook a validation exercise by seeking input from our clinical expert on these costs.

The ERG are aware that clinical practice varies across centres for liver transplantation patients. However, Novartis obtained their resource use data through 1 centre and it is therefore important to validate these against different sources. In general, clinical opinion sought by the ERG disagreed with some of the resource data reported in the submission.

Our clinical advisor indicated that overall, GP visits are unlikely to occur as frequently as most of these patients will be mainly managed in secondary care and not in primary care. So for example, for the SPT state it is likely that the patients would need an additional appointment in the hospital (despite of the possible GP visit). For the ASRR state, the company indicate 6 visits to the GP, however these are far more likely to be 6 hospital outpatient visits as GPs are not involved in the management of acute steroid resistant rejection cases. It was also mentioned that for the AR state, 3 hospital outpatient visits were likely to be necessary, instead of the 1 visit assumed by the Novartis model.

With regards to tests, the ERG's clinical advisor indicated that there might be an underestimation of the number of tests required in some health states. For example, in the SPT state, patients are also likely to require a blood test to check immunosuppression trough levels.

Furthermore, it was noted that a liver biopsy requires 1 day in the hospital. Novartis present a cost of £514 for liver biopsy taken form the NHS 2013 reference costs, however, the ERG searched the code and description of the procedure indicated by the company but the costs associated with this do not match Novartis costs (we found a cost estimate of £506 for short stay and a cost of £236 for outpatient procedures). Therefore the ERG are not clear where the cost was taken from and if it includes hospital stay or not. Finally our clinical advisor indicated that for the MCR state all the resource data seemed too high to manage a condition which is asymptomatic and doesn't require treatment and that the number of tests required in the ASRR state were likely to be 8, instead of 6.

Superseded
See Erratum

The ERG found a reporting mistake related to the number of hospital outpatient visits in the ASRR state. The submission reports this value as 14 days, however, in the excel model this value is 1 day. Given this is the state of graft loss, it seems reasonable that this value is 14 days.

Given that patients in the MCR state do not suffer from any QoL loss (compared to patients in the SPT state) the difference in costs between these health sates is not clear. The cost associated with the MCR health state is £640, whilst the cost associated with the SPT state is £73. Given that patients can stay in the MCR state for several years, and that the company has not provided any reason for the resource use to be threefold that of the SPT state, it would appear reasonable to assume instead a similar resource use in these health states.

For the renal sub-model, Novartis reported NICE CG182 and Kerr et al, 2012 as the two sources used to obtain resource use data. Despite Kerr et al, 2012 not being reported in the literature review, the ERG traced the source back to the CG182 document, which used this study. The company also did not report NICE CG73, which in the economic model features as a resource data source used several times.

The ERG could not verify all the resource use data as CG73 doesn't seem to be available online (as it has been replaced by CG182).

The frequency of GP visits was obtained from Kerr et al, 2012, however this study only looked into CKD stage 3, 4 and 5. Nonetheless, Novartis assumed the same frequency of GP visits for CKD stage 1 and 2 as the ones reported in the study for stages 3, 4 and 5. It seems unlikely that patients on CKD stage 1 would require the same monitoring as patients on CKD stages 4 or 5.

In the excel model, a patient incurs the cost of transplantation on the same cycle that renal transplantation occurs. On that same cycle, the post-transplantation ongoing care costs accrue to this patient. The patient keeps accruing this cost every cycle until death (in addition to the cost associated with whatever CKD stage the patient is on).

Table 48 provides an example taken from the excel model for one simulated patient. It seems extremely unlikely that patients would accrue such high annual post-transplantation costs for nearly 20 years (the source of the cost, Kerr et al 2012 does not provide details about what is considered in this cost).

Table 48. Post-transplantation ongoing care costs in the renal model

Health state	Time	Cost accrued in the model
CKD stage 5	14 cycles (4 years)	CKD stage 5 costs
No CKD (after transplantation)	52 cycles (13 years)	Transplantation cost (one off) + Post-transplant ongoing care costs
CKD stage 1/2	19 cycles (5 years)	CKD stage 1/2 costs + Post-transplant ongoing care costs
Death	Remaining cycles	No costs

Finally, all of the costs presented for the CKD health states were wrong in the submission (when compared with the economic model). This was is a misreporting problem, as the resource use and the unit costs in the submission are correct and match the ones in the excel model, used to derive the total costs.

Intervention and comparators' costs

Novartis have used the Prograf brand price to cost the use of TAC in the model, however there is no justification given for this decision. The ERG looked at the prices for different brands of TAC, and Prograf is the most expensive one (£1.61 per mg), with Vivadex being the least expensive (£1.2 per mg). As there is no apparent justification for choosing Prograf over any other brand used in the UK, the ERG took the average across all brand prices and derived an estimate of £1.3 per mg. Ideally the company would have used a weighted average, according to market shares.

No on-going monitoring costs were included as it was assumed that that these were included in the consultations costs (page 191 of Novartis submission)

The company states that dose assumptions for EVR+rTAC were taken from H2304, while dosages for other drugs were based on the BNF. These are presented in Table 49. The ERG could not find the TAC doses used in H2304 in the CSRs, as TAC prescription is usually reported as trough levels to be achieved and not dose. This was also the case for most studies in the NMA, where TAC was prescribed according to target trough levels.

Clinical opinion sought by the ERG advised that even though it is reasonable to assume that drug doses will remain overall constant after 1 year, the primary focus is to keep TAC doses as

low as possible for a specific patient, hence some of these doses were considered slightly high to be maintained for the rest of the patients' lives. The ERG's clinical advisor also mentioned that the doses of oral MMF seem high after 1 year, and that 1000 or 2000 (and not 3000) mg would reflect UK clinical practice more accurately for the 1 year post transplantation period.

Table 49. Mean daily dose of immunosuppression assumed in the model

	Starting dose	Month 3	Month 6	Month 9	Month 12+
ARM1 Everolimus regimen					
Everolimus	2.0 mg/day	2.0 mg/day	2.0 mg/day	2.0 mg/day	2.0 mg/day
Tacrolimus	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Corticosteroid	15.0 mg/day	7.5 mg/day	0.0 mg/day	0.0 mg/day	0.0 mg/day
ARM2 Azathioprine regimen					
Azathioprine	1.0 mg/kg/day	1.0 mg/kg/day	1.0 mg/kg/day	1.0 mg/kg/day	1.0 mg/kg/day
Tacrolimus	0.12 mg/kg/day	0.09 mg/kg/day	0.08 mg/kg/day	0.07 mg/kg/day	0.07 mg/kg/day
Corticosteroid	15.0 mg/day	7.5 mg/day	0.0 mg/day	0.0 mg/day	0.0 mg/day
ARM3 Mycophenolate mofetil regimen					
Mycophenolate mofetil intravenous	2000 mg/day	0 mg/day	0 mg/day	0 mg/day	0 mg/day
Mycophenolate mofetil oral	3000 mg/day	3000 mg/day	3000 mg/day	3000 mg/day	3000 mg/day
Tacrolimus	0.12 mg/kg/day	0.09 mg/kg/day	0.08 mg/kg/day	0.07 mg/kg/day	0.07 mg/kg/day
Corticosteroid	15.0 mg/day	7.5 mg/day	0.0 mg/day	0.0 mg/day	0.0 mg/day
ARM4 dummy regimen					
Dummy drug	0.0 mg/day	0.0 mg/day	0.0 mg/day	0.0 mg/day	0.0 mg/day
Dummy drug	0.0 mg/day	0.0 mg/day	0.0 mg/day	0.0 mg/day	0.0 mg/day
Corticosteroid	0.0 mg/day	0.0 mg/day	0.0 mg/day	0.0 mg/day	0.0 mg/day

5.2.4 Adverse events

The ERG are generally satisfied with the estimation of AE occurrence and the QoL data used to estimate these in the model. With regards to cost data, the ERG are not clear about the choice of the source for the cost of diabetes. This was the Davidson et al., 2003 study, which has not been referenced in the systematic literature search. Furthermore, no details are provided as to

which resources are considered for the treatment costs of diabetes, which are reported to be £1,840.

The recurrence of HCV and HCC was not considered in the model as Novartis state that clinical literature and consultation with clinical experts indicated that there is no effect of the immunosuppressive regimen on these variables.

From a conceptual point of view, the ERG are not satisfied that the costs and utility losses associated with EVR-related AE have been included for the 3 months in the first model cycle. Given that EVR therapy only starts 1 month after surgery (and drug costs were considered for 2 months only), the AE associated with the drug should have been considered only for 2 months as well.

5.2.5 Sensitivity analysis

Deterministic sensitivity analysis was not undertaken in Novartis' submission. This is unfortunate as several model parameters were based on clinical assumptions (especially resource use). It would be useful to understand what the key drivers of the economic model are, especially considering the high ICERs presented.

Scenario analysis was run where model structure was changed (removal of different health states).

Novartis ran a PSA using 1,000 probabilistic simulations. In two different occasions the company reported that 5,000 patients where used in the simulation and that 1,000 were used in the simulation (as using 5,000 patients supposedly produced the same results). These values are reported to have been chosen to balance the model stability with practical run time.

Nonetheless, as previously mentioned by the ERG, it is likely that the base case simulation model is not stable with 10,000 individual patients (i.e. simulations). Therefore, running the model with 1,000 is even more likely to generate unreliable estimates. It is nonetheless surprising that Novartis report that running the model with 1,000 patients produced the same results as running it with 5,0000.

5.3 Results included in company's submission

This section presents a summary of the results of Novartis' model.

5.3.1 Deterministic results

- Base case

The ERG present Novartis' base case results in Table 50. The ERG find the original results table provided by Novartis slightly confusing. EVR+rTAC presents a cost per QALY gained of

£187,842 compared with AZA+TAC and a cost per QALY gained of £110,797 when compared with MMF+TAC.

EVR+rTAC presents higher costs than the comparators over the 80 years of analysis. The incremental gain in QALYs is 0.188 and 0.343 compared with AZA+TAC and MMF+TAC, respectively. AZA+TAC seems to be the cheaper option amongst the 3 treatments, and the one with the smallest difference in terms of QALY loss compared with EVR+rTAC.

Table 50. Base case ICERs

Cost-effectiveness results per patient	EVR+rTAC (1)	AZA+TAC (2)	MMF +TAC (3)	Incremental value (1-2)	Incremental value (1-3)
Total costs £	140,750	105,529	102,746	**£35,221**	**£38,004**
QALYs	4.210	4.022	3.867	**0.188**	**0.343**
ICER				**£187,842**	**£110,797**

During the clarification process the ERG have requested that Novartis provided some rationale to explain the high ICER estimates. The company replied that all comparators have similarly high efficacy with regard to prevention of rejection-related events and that EVR is a branded product being compared to generic comparators, which may partially explain the high ICERs presented.

Novartis also presented the average number of cycles spent in the different states of the economic model. The ERG noted a serious logical error in the model when analysing these data. Table 51 presents these.

There are a total of 320 cycles in the economic model (ran for 80 years). As the hepatic and the renal sub-model take place simultaneously and there is no relationship between the models whatsoever (except death), the average amount of time spent (i.e. time alive) in one model needs to equal the average amount of time spent in the other. Since patients don't "leave one model to go into the other", anytime patients leave either models is because they have died.

Looking at Table 51 below, it can be observed that patients spent, on average, 41 cycles in the EVR+rTAC arm of the hepatic model. The rest of the cycles (i.e 320 – 41 = 279) patients were dead. However, when looking at the renal model, patients only spent 31 cycles alive in the model.

Because these average results were obtained through the original simulation model, the ERG cannot trace the values back. Looking at the model formulae, the ERG could not identify the problem. However the average number of cycles spent in the dead state is the same for both the hepatic and the renal sub-model, which indicates that the problem is not related to how patient death was implemented in the simulation model.

As the number of cycles in the dead state seem to be correct, the ERG believe the problem is in the renal sub-model. Looking at the number of cycles spent in the dead state in the EVR+rTAC arm in Table 51, these amount to 279. If we add the number of cycles spent alive in the hepatic model, 41, these add up to 320 as they should. However, when we add the number of cycles spent alive in the renal model, 31, to the 279 cycles, this only adds up to 310. So there are 10 cycles “missing” from the renal sub-model. Clearly this logical error reflects a problem in the model formulae and/or structure.

Table 51. Average time spent in different health states of the economic model

Cycles in state	Everolimus Regimen	Azathioprine Regimen	MMF Regimen
Stable post-transplant	33.012	31.735	31.327
Acute Rejection	0.785	0.771	0.363
Steroid resistant acute rejection	0.014	0.015	0.006
Mild chronic rejection	6.777	6.571	6.974
Severe rejection	0.005	0.004	0.002
Stable post re-transplant	0.002	0.002	0.001
Total	**40.595**	**39.097**	**38.673**
No CKD	0.000	0.000	0.000
CKD1to2	5.278	4.848	4.753
CKD3	4.453	4.332	4.000
CKD4	14.421	14.328	14.129
CKD5	2.729	3.164	3.128
No CKD post-transplant	4.271	4.868	4.799
Total	**31.152**	**31.541**	**30.809**
AE disutility	-0.353	-0.424	-1.551
Number of cycles in dead state	**279.405**	**280.903**	**281.327**
Number of total cycles = 320			
Years alive	10.1	9.8	9.7
Years dead	69.9	70.2	70.3

- ## Scenario analysis

Novartis ran 4 scenario analysis. The ERG could not find the scenario analysis in the excel model, therefore, we report the results presented in the submission but we note that these analyses have not been validated or found by the ERG in the economic model. This is particularly unsettling given that the face validity of the results is somewhat questionable.

Novartis have run the following scenario analysis:

1. Elimination of the Mild Chronic Rejection (MCR) state from the core hepatic model. Justification provided:

   - *This was performed as it is assumed that a patient who enters this state does not do so on account of their immunosuppressive regimen but rather it has a non-treatment specific underlying cause.*

2. Re-transplantation was removed from the analysis. Justification provided:

   - *It was expected that by removing these options for future transplants, patients would persist longer in the end states in both models and, because these states carry a high mortality rate, it was expected that the life years lived and the QALYs in each treatment arm would fall*.

3. The core cost-effectiveness model (hepatic rejection) was run without the renal sub-model. Justification provided:

   - *By running the main (hepatic rejection) model only, it is expected to show minimal difference in life years lived and QALYs gained between treatment arms due to the good management of rejection. Such a result validates the hypothesis that the real value of this exercise is the renal sparing affect that everolimus with reduced dose tacrolimus and corticosteroids offers.*

4. The baseline eGFR was varied from 81 mL/ min per 1.73 m^2 to 60mL/min per 1.73 m^2.

   - *One difference highlighted between the H2304 trial and UK clinical practice was that the trial contained non-UK patients, dominated by US patients.*
   - *A discussion with a clinical expert revealed the profile of UK candidates for immunosuppressive treatment is usually a CKD stage lower than the US and the baseliner range would be more realistically around 55-65 mL/min per 1.73 m^2 post-transplant.*
   - *It is expected that lowering the eGFR used at the model's start will speed up the transitions to CKD Stage 5 and consequential renal transplantation following this.*

Results for scenario 1, 2 and 4 are presented in Table 52, Table 53 and Table 54, respectively.

The ERG do not find scenario 3 particularly insightful, hence it is not reported here. The reported ICER for EVR+rTAC versus MMF+TAC was £312,279 cost per QALYs.

The results of the analysis undertaken in scenario 1 are quite surprising. There does not seem to be consistency on the impact of removing the MCR state from the model when the base case results are compared in the EVR+rTAC and AZA+TAC with the MMF+TAC treatment arm. While the EVR+rTAC and AZA+TAC costs decreased and the total QALYs increased (compared to the base case), for MMF+TAC the total costs increased as did the total QALYs, compared with the MMF+TAC base case. Also, the ICER for EVR+rTAC compared with MMF+TAC doubled, going from £110,797 to £227,528. The ICER comparing EVR+rTAC with AZA+TAC decreased, going from £187,842 to £176,410 in scenario 1.

What is also remarkable is that only 1% of patients can move to the MCR state after 1 year in the SPT state in all treatment arms of the model. The state has been considered asymptomatic, therefore there is no QoL loss considered when patients move from the SPT to the MCR state. Nonetheless, removing this state from the model doubled the base case ICER of EVR+rTAC compared with MMF+TAC but lead to a decrease in the ICER comparing EVR+rTAC with AZA+TAC.

Superseded
See Erratum

These differences can only be explained trough the difference in costs associated with the MCR state, which are considerably higher than the ones in the SPT state. However, the ERG cannot find a plausible reason why the impact of removing the MCR state would increase the ICER comparing EVR+rTAC with MMF+TAC and decrease the ICER comparing the intervention with AZA+TAC.

As previously mentioned, the ERG question the clinical validity of the MCR state but also the assumptions made by the company with regards to the resource use in this state. This is further explored in the next section of the report.

Table 52. Scenario 1 run by Novartis

Cost-effectiveness results per patient	EVR+rTAC (1)	AZA+TAC (2)	MMF +TAC (3)	Incremental value (1-2)[4]	Incremental value (1-3)
Total costs £	140,189	104,907	102,856	**£35,282**	**£37,333**
QALYs	4.25	4.05	4.09	**0.2**	**0.16**
ICER				**£176,410**	**£227,528**

Removing the re-transplantation option from both models, again, lead to different results with regards to the final ICERs. Compared with the base case ICERs, the ICER for EVR+rTAC versus AZA+TAC decreased whilst the ICER for EVR+rTAC versus MMF+TAC increased.

The total QALYs decreased across all treatment arms, as the company hypothesised.

[4] These values were estimated by the ERG. As the total costs and QALYs were not provided in the excel model for the scenario analysis and Novartis did not report the incremental analysis between EVR+rTAC and AZA+TAC, the ERG used the values presented in the tables to estimate incremental cost-effectiveness ratios. Differences in ICERs may be due to rounding.

Table 53. Scenario 2 run by Novartis

Cost-effectiveness results per patient	EVR+rTAC (1)	AZA+TAC (2)	MMF +TAC (3)	Incremental value (1-2)[4]	Incremental value (1-3)
Total costs £	149,418	122,468	119,924	**£26,950**	**£29,494**
QALYs	3.69	3.46	3.45	**0.23**	**0.24**
ICER				**£117,174**	**£117,285**

For scenario 4, Novartis decreased the baseline eGFR level from 81 mL/ min per 1.73 m^2 to 60mL/min per 1.73 m^2. The reason provided for this was that the eGFR baseline levels in H2304 might be higher than the ones usually observed in UK clinical practice.

Similarly to the other scenario analysis, there was not consistency in the impact on the final ICERs across treatment arms. Compared with the base case ICERs, the ICER for EVR+rTAC versus AZA+TAC decreased whilst the ICER for EVR+rTAC versus MMF+TAC increased.

Table 54. Scenario 4 run by Novartis

Cost-effectiveness results per patient	EVR+rTAC (1)	AZA+TAC (2)	MMF +TAC (3)	Incremental value (1-2)[4]	Incremental value (1-3)
Total costs £	160,845	131,392	129,448	**£29,453**	**£31,397**
QALYs	3.79	3.63	3.62	**0.16**	**0.17**
ICER				**£184,081**	**£184,372**

5.3.2 Probabilistic sensitivity analysis results

Novartis ran a PSA using 1,000 probabilistic simulations. In two different occasions the company reported that 5,000 patients where used in the simulation and that 1,000 were used in the simulation (as using 5,000 patients supposedly produced the same results as 1,000 simulations).

As previously mentioned by the ERG, it is likely that the base case simulation model is not stable with 10,000 patients (i.e. simulations). Therefore, running the model with 1,000 is even more likely to generate unreliable estimates.

The results for EVR+rTAC versus AZA+TAC showed an incremental cost per QALY gained of £184,714 (compared to the base case ICER of £187,842).

The results for EVR+rTAC versus MMF+TAC showed an incremental cost per QALY gained of £105,526 (compared with the base case ICER of £110,797).

Again, the ERG note that these results are based on a small number of simulations and are likely to lack robustness.

6.0 Additional work undertaken by the ERG

In this section we explore the implications of some of the errors found in Novartis' model. The ERG were faced with two challenges with regards to the best approach to undertake additional analysis:

1. The fact that the base case analysis is extremely unstable and unreliable.
2. The time constraints associated with the review process and the computational burden of the patient simulation model

The ERG ran two iterations of the base case model as submitted by the company, using the number of recommended simulations and **without any changes,** to test the model stability with regards to ICER results. The ERG found a considerable variation in the ICERs reported, especially in the case of the EVR+rTAC versus MMF+TAC ICERs, which varied between £110,797 and £120,651 (nearly a 9% change).

As the base case results are so unstable, if we were to run additional analysis with different input values, the final impact on the ICER would not be identifiable as it would be confounded with quasi-random simulation error and as such, it would not produce a reliable or valid ICER.

Additionally, it took the ERG around 1 hour to run each simulation of the model. As the ERG have identified a considerable number of problems with model inputs, we identified at least 7 scenarios of potential relevance for investigation, which would have amounted to a significant time burden.

The ERG therefore considered more relevant to try to shed some light on the reason why the ICERs vary so much when running the simulation model several times. This is explored in Section 6.1.

6.1 Correction for errors in Novartis' model

The errors identified by the ERG that could potentially lead to alterations in the model results are:

- Given that patients in the MCR state do not suffer from any QoL loss (compared to patients in the SPT state) the difference in costs between these health sates is not clear. The cost associated with the MCR health state is £640, whilst the cost associated with the SPT state is £73. Given that patients can stay in the MCR state for several years, and that the company has not provided any reason for the resource use to be threefold that of the SPT state, it would appear reasonable to assume a similar resource use in these health states. Therefore we would have liked to run an analysis where the costs for the MCR state are the same as for the SPT state.

- Novartis have used the Prograf brand price to cost the use of TAC in the model. However there was no justification given for this decision. The ERG looked at the prices

for different brands of TAC, and Prograf is the most expensive one (£1.61 per mg), with Vivadex being the least expensive (£1.2 per mg). As there is no apparent justification for choosing Prograf over any other brand used in the UK, the ERG took the average across all brands and got a price of £1.3 per mg. We would have liked to explore the impact of this change in the final ICER.

- The ERG found a reporting mistake related to the number of hospital outpatient visits in the ASRR state. The submission reports this value as 14 days, however, in the excel model this value is 1 day. The ERG would have liked to correct this.

- From a conceptual point of view, the ERG are not satisfied that the costs and utility losses associated with EVR-related AE have been included for the 3 months in the first model cycle. Given that EVR therapy only starts 1 month after surgery (and drug costs were considered for 2 months only), the AE associated with the drug should have been considered only for 2 months as well.

- Novartis use baseline proportions as progression rates of renal disease in the renal sub-model. The ERG believe they should have used the values reported in Table 45 instead.

- The ERG found a mistake in the reporting of the SCR utility value. In the submission, a utility value of 0.43 is reported for this health state; however, the economic model uses a utility value of 0.53.

- The estimate used for the decrease in renal function at 12 months for the MMF and AZA arms of the economic model was based on the decrease in renal function for "standard" TAC. The ERG believe that the MMF+ "reduced" TAC arm estimate should be used in the MMF arm of the economic model instead of the standard TAC estimate because the MMF+ "reduced" TAC is actually closer to a UK MMF+ standard TAC arm if we take the H2304 study (and UK clinical practice) as reference. The same value could be used for the AZA arm of the model as the renal dysfunction is determined by the levels of TAC (and is not dependent on the concomitant drug).

The ERG tried to determine the cause for instability in the model results. Therefore we have "fixed" the baseline characteristics of patients (by taking their mean values) in the simulation model instead of allowing these values to vary each simulation according to a distribution. The point of this analysis was to understand if the variation in results is generated by the simulated patient characteristics or if this is attributable to other problems in the model. The results of this exercise originated ICERs for EVR+rTAC versus AZA+TAC ranging **from dominant to £797,558** and ICERs for EVR+rTAC ranging **from £431,348 to £582,668**.

This suggests that the model instability cannot be solved by fixing patient baseline characteristics as mean estimates. Furthermore, the variance in results for EVR+rTAC compared with AZA+TAC seems greater than the one obtained for EVR+rTAC versus MMF+TAC.

7.0 Summary of clinical and cost-effectiveness issues

The ERG are extremely concerned with the **lack of stability in the base case ICERs**. The ERG ran two iterations of the base case model as submitted by the company, using the number of recommended simulations and **without any changes,** to test the model stability with regards to ICER results. The ERG found a considerable variation in the ICERs reported, especially in the case of the ICERs for EVR+rTAC versus MMF+TAC, which varied between £110,797 and £120,651 (nearly a 9% change).

The ERG are overall concerned with the **validity of the NMA** undertaken by Novartis. The approach undertaken raises the following concerns:

- Even though WinBUGS codes were included in the submission, these did not provide any relevant information. Whilst the codes presented reflect a standard NMA approach, these seem to have been either taken from other submission or from a theoretical exercise, as the data included in the codes do not relate to the submission data in any way. This poses a major limitation as the ERG could not verify which data were used for the analysis of specific outcomes. This represents a major concern as the ERG are not clear which studies have been included in the NMA analysis for the tBPAR outcome, due to lack of clarity and transparency in the submission.

- Overall, the ERG question the validity of the NMA results for the renal outcomes. The allocation of different studies' treatment arms to the reduced and standard TAC categories is inconsistent and misleading:

    - The ERG are deeply concerned with the fact that the MMF arms of the trials were considered by the company to include reduced doses of TAC. The trough levels of TAC in the "reduced" arms of Neuberger et al, 2009 and Boudjema, et al 2011 studies are similar to the trough levels of standard TAC in H2304. The average trough levels in the standard TAC arms in both studies are generally higher than the trough levels in the standard TAC arm of H2304. This means that the "reduced" TAC arms in these studies are in fact standard TAC arms. And even if these are considered standard TAC levels, they are still higher than the ones verified in UK clinical practice.

    - The ERG are equally concerned with the fact that the company decided to consider the levels of TAC in the Porayko et al., 1994 study as "standard" TAC. The TAC trough levels were maintained below 1.0 ng/mL for the entire year (beginning 4 weeks after randomization), which actually **represents the lower trough levels of TAC amongst all the of NMA studies (including H2304).** So in reality the study is comparing AZA+ ciclosporin with reduced TAC. The same is applicable to the McDiarmid et al, 1993 study, where TAC trough levels were maintained below 2 ng/mL.

As the standard TAC connector across the NMA studies is so heterogeneous (note for example that in the Boudjema et al, 2009 study the trough level of TAC in the standard TAC arm was close to 12 ng/mL for several months during the first year) the NMA results are likely to lack robustness.

The ERG are overall concerned with the **model structure and assumptions** used by Novartis. The approach undertaken raises the following concerns:

- The ERG are not convinced that a patient simulation state-transition model is required. The company has not reported sufficient details or a solid justification for their approach. It seems that a traditional cohort state-transition model would be appropriate in this case and more transparent that the model adopted by the company.

- The ERG believe that more emphasis should have been placed on the renal component of the economic model. More specifically, interaction between the two model components (hepatic and renal) should have been considered as it is known that not only immunosuppressive therapy for liver post-transplantation has an impact on renal functioning but also, equally important, that renal functioning has an impact on graft survival.

- The ERG consider the cycle length too large to reliably capture clinically important events. We also question the clinical plausibility of the MCR state and especially the high costs associated with this state. Removing this state from the economic analysis led to an increase in the ICER for EVR+rTAC compared with MMF+TAC, going from £110,797 to £227,528, whilst the ICER comparing EVR+rTAC with AZA+TAC decreased, going from £187,842 to £176,410.

- Overall survival is underestimated in the economic model.

- The way the simulation was run in the model produces clinical implausible results. Due to the random generation process used to estimate the baseline eGFR levels, some patients start the model with negative levels of eGFR, which is not realistic. These patients can spend years in the CKD 5 stage of the renal with a negative eGFR level. Clearly this is not a plausible representation of clinical reality.

- We have identified errors in the formulae allocating patients to different health states in the model which:

    - Are likely to have been avoided if cohort state-transition model had been used.
    - Could not be corrected by the ERG due to the computational burden of running the simulation model.
    - Are likely to bias the results in favour of EVR+rTAC in the model.

The ERG are worried with the **clinical inputs** used by Novartis. The ERG found that some of the parameters being used in the economic model differ from the ones reported in the

submission and that some assumptions made with regards to resource use and drug effectiveness on renal outcomes should be further tested and validated.

Finally, the fact that the EVR+rTAC, the MM+TAC and AZA+TAC arms are not reflective of clinical reality in the UK in terms of TAC levels is worrying as it questions the validity and the applicability of the economic analysis within the NHS context.

For all the reasons provided, the ERG lack confidence in the final ICER presented and in the company analyses and therefore cannot make any predictions regarding the correct final ICERs.

8.0 References

Abad J, Garcia J, Sansi A, Eizaguirre E, Vargas L, Navarro R, et al. Impact of renal retransplantation on graft and recipient survival. Transplant. 2011;64(4):363-70.

Annemans L, Moeremans K, Mutimer D, Schneeberger H, Milligan D, Kubin M. Modeling costs and cost-effectiveness of different CMV management strategies in liver transplant recipients as a support for current and future decision making. Value in Health. 2002;5(4):348-58.

Aydogan C, Sevmis S, Aktas S, Karakayali H, Demirhan B, Haberal M. Steroid-resistant acute rejections after liver transplant. Exp Clin Transplant. 2010;8(2):172-7.

Aylin P, Alexandrescu R, Jen MH, Mayer EK, Bottle A. Day of week of procedure and 30 day mortality for elective surgery: retrospective analysis of hospital episode statistics. BMJ. 2013;346:f2424.

Barber K, Blackwell J, Collett D, Neuberger J. Life expectancy of adult liver allograft recipients in the UK. Gut 2007;56(2):279-82

Boudjema K, Camus C, Saliba F, Calmus Y, Salame E, Pageaux G, et al. Reduced-dose tacrolimus with mycophenolate mofetil vs. standard-dose tacrolimus in liver transplantation: a randomized study. AmJTransplant. 2011;11(5):965-76.

Briggs, A. H., Claxton, K., & Sculpher, M. J. (2006). Decision modelling for health economic evaluation. Oxford university press

Burroughs AK, Blake J, Thorne S, Else M, Rolles K. Comparative hospital costs of liver transplantation and the treatment of complications of cirrhosis. European Journal of Gastroenterology and Hepatology. 1992;4(2):123-8.

Caro, J. J., Briggs, A. H., & Siebert, U. (2011). DRAFT—Modeling Good Research Practices—Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. ISPOR Task Force Report.

Charlton M, Wall W, Ojo A, Gines P, A. T, Shihab F, et al. Report of the first internation liiver transplantatio society expert panel consensus conference on renal insufficiency in liver transplantation. Liver Transplantation. 2009;15:S1-S34.

Cillo U, Vitale A, Salizzoni M, Coledan M, Rossi G, Baccarani U, et al. Early Introduction of everolimus in de novo liver transplantation: A multicenter randomized clinical trial (EPOCAL). Journal of Hepatology. 2013;58:S63-S227.

Curtis L. Unit Costs of Health & Social Care 2013. 2013.

Davidson J, Wilkinson A, Dantal J, Dotta F, Haller H, Hernandez D, et al. New-onset diabetes after transplantation: 2003 International consensus guidelines. Proceedings of an international expert panel meeting. Barcelona, Spain, 19 February 2003. Transplantation. 2003;75(10 Suppl):SS3-24.

Davis S, Stevenson M, Tappenden P, Wailoo A. NICE DSU Technical Support Document 15: Cost-effectiveness modelling using patient-level simulation. 2014

de Ruvo N, Riva G, di Benedetto F, Tarantino G, Potenza L, Ballarin R, et al. Thymoglobuline pre-treatment and minimization of immunosuppression with Advagraf in liver transplantation allows increase of t-reg FoxP3 + cells and guidance of tolerogenesis. Hepato-Pancreato-Biliary. 2012;14(2):107-287.

de Simone P, Jones R, Kohli V, Gedaly R, Jeffrey G, Cantisani G, et al., editors. Everolimus facilitated reduction of tacrolimus maintains comparable efficacy and better renal function versus standard tacrolimus in de novo liver transplant recipients. ILTS; 2013.

De Simone P, Metselaar HJ, Fischer L, Dumortier J, Boudjema K, Hardwigsen J, et al. Conversion from a calcineurin inhibitor to everolimus therapy in maintenance liver transplant recipients: a prospective, randomized, multicenter trial. Liver Transpl. 2009;15(10):1262-9.

De Simone P, Nevens F, De Carlis L, Metselaar HJ, Beckebaum S, Saliba F, et al. Everolimus with reduced tacrolimus improves renal function in de novo liver transplant recipients: a randomized controlled trial. Am J Transplant. 2012;12(11):3008-20.

Dias, S., Welton, N. J., Sutton, A. J., & Ades, A. E. (2011). NICE DSU Technical Support Document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. National Institute for Health and Clinical Excellence, London, UK.

Dmitrewski J, Ayres S, Gunson BK, Buist LJ, Buckels JA, McMaster P, et al. Steroid withdrawal 3 months after liver transplantation--does FK 506 confer any advantage over cyclosporin? Transplant International. 1994;7 Suppl 1:S85-S7.

Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. J., & Stoddart, G. L. (2005). Methods for the economic evaluation of health care programmes.OUP Catalogue.

Fischer L, Klempnauer J, Beckebaum S, Metselaar H, Neuhaus P, Schemmer P, et al. A Randomized, Controlled Study to Assess the Conversion From Calcineurin-Inhibitors to Everolimus After Liver Transplantation—PROTECT. American Journal of Transplantation. 2012;12:1855-186.

Foxton M, Al-Freah M, Portal A, Sizer E, Bernal W, Auzinger G, et al. Increased model for end stage liver disease score at the time of liver transplant results in prolonged hospitalization and pverall intensive care unit costs. Live Transplantation. 2010;16:668-77

Guidelines for Liver Transplantation for Patients with Non-Alcoholic Steatohepatitis. British Society of Gastroenterology. Available from http://www.bsg.org.uk/

Hexham JM, Wang Y, Lopez PM, Jiang H, Junge G. A 24 month, multi-center, open-label, randomized, controlled study to evaluate the efficacy and safety of concentration-controlled everolimus to eliminate or to reduce tacrolimus compared to tacrolimus in de novo liver transplant recipients. H2304 Clinical Study Reports 2011.

Hirschfield GM, Gibbs P, Griffiths WJ. Adult liver transplantation: what non-specialists need to know. BMJ. 2009.

Jain A, Kashyap R, Kramer D, Dodson F, Hamad I, Starzl TE, et al. Prospective randomized trial of tacrolimus and prednisone versus tacrolimus, prednisone, and mycophenolate mofetil: complete report on 350 primary adult liver transplantations. Transplant Proceedings. 2001;33(1-2):1342-4.

Junge G, Dumortier T, Fischer L, Kohler S, Fung J. Everolimus (RAD001) treatment effect in liver transplantation-a rational approach to quantify the mTOR effect. American Journal of Transplantation. 2013;13:504-5

Junge G, Dumortier T, Schwende H, Fung J. mTOR Inhibition in Liver Transplantation: How to Dose for Effective/Safe CNI Reduction? Transplantation Proceedings. 2013;45:1979-80.

Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Moller J. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. Med Decis Making. 2012;32(5):701-11.
Keith DS, Nichols GA, Gullion CM, Brown JB, Smith DH. Longitudinal follow-up and outcomes among a population with chronic kidney disease in a large managed care organization. Arch Intern Med. 2004;164(6):659-63.

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

La Mattina JC, Foley DP, Mezrich JD, Fernandez LA, Vidyasagar V, D'Alessandro AM, et al. Chronic kidney disease stage progression in liver transplant recipients. Clin J Am Soc Nephrol. 2011;6(8):1851-7.

Liem YS, Bosch JL, Hunink MG. Preference-based quality of life of patients on renal replacement therapy: a systematic review and meta-analysis. Value Health. 2008;11(4):733-41.

Lopez PM, Hexham JD, Dong G, Kohler S, Ng J, Junge G. A 24 month, multi-center, open-label, randomized, controlled study to evaluate the efficacy and safety of concentration-controlled everolimus to eliminate or to reduce tacrolimus compared to tacrolimus in de novoliver transplant recipients (CSR1). H2304 Clinical Study Reports 2013.

McDiarmid SV, Colonna JO, Shaked A, Ament ME, Busuttil RW. A comparison of renal function in cyclosporine- and FK-506-treated patients after primary orthotopic liver transplantation. Transplantation. 1993;56(4):847-53.

McGuire BM, Rosenthal P, Brown CC, Busch AM, Calcatera SM, Claria RS, et al. Long-term management of the liver transplant patient: recommendations for the primary care doctor. Am J Transplant. 2009;9(9):1988-2003.

Meier-Kriesche HU, Li S, Gruessner RW. Immunosuppression: evolution in practice and trends, 1994-2004. AmJTransplant. 2006;6(5-2):1111-31.

Neil DA, Adams DH, Gunson B, Hubscher SG. Is chronic rejection of liver transplants different from graft arteriosclerosis of kidney and heart transplants? Transplant Proc. 1997;29(6):2539-40.

Neri L, McEwan P, Sennfalt K, Baboolal K. Characterizing the relationship between health utility and renal function after kidney transplantation in UK and US: a cross-sectional study. Health Qual Life Outcomes. 2012;10:139.

Neuberger JM, Mamelok RD, Neuhaus P, Pirenne J, Samuel D, Isoniemi H, et al. Delayed introduction of reduced-dose tacrolimus, and renal function in liver transplantation: the 'ReSpECT' study. Am J Transplant. 2009;9(2):327-36.

NHS Blood and Transplant [Internet]. Available from http://www.organdonation.nhs.uk/

Novartis, National Institute for Health and Care Excellence, Single technology appraisa,l Everolimus (Certican®) for preventing organ rejection in liver transplantation, Specification for manufacturer/sponsor submission of evidence. 24th September 2014

Office for National statistics 2012. Available from http://www.ons.gov.uk

Organ Donation. Kidney Activity. 2014.

Organ Donation. Kidney Activity. 2014.

Philips Z, Bojke L, Sculpher M, Claxton K, Golder S: Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. PharmacoEconomics 2006, 24(4):355-371

Philips, Z., Ginnelly, L., Sculpher, M., Claxton, K., Golder, S., Riemsma, R., ... & Glanville, J. (2004). Review of guidelines for good practice in decision-analytic modelling in health technology assessment.
Porayko MK, Textor SC, Krom RA, Hay JE, Gores GJ, Richards TM, et al. Nephrotoxic effects of primary immunosuppression with FK-506 and cyclosporine regimens after liver transplantation. Mayo Clin Proc. 1994;69(2):105-11.

Prograf SPC.[Internet] Available from https://www.medicines.org.uk/

Ratcliffe J, Longworth L, Young T, Bryan S, Burroughs A, Buxton M. Assessing Health-Related Quality of Life Pre– and Post– Liver Transplantation: A Prospective Multicenter Study. Liver Transplantation. 2002;8:263-70.

Saliba F, Kaiser GM, de Carlis L. Everolimus-based immunosuppression provides superior renal function and comparable efficacy versus standard tacrolimus in de novo liver transplantation: 3-year results from H2304 study extension. Journal of Hepatology 2014:S362.

Samuel D, Bismuth H, Boillot O, Ducerf C, Baulieux J, Gugenheim J, et al. Tacrolimus (FK506)-based dual versus triple therapy following liver transplantation. Transplant Proc. 1998;30(4):1394-6.

Schlitt HJ, Lehner F, Neuhaus P, Schmidt J, Becker T, Hack MA, et al. The protect study: Preservation of renal function in liver transplant recipients with certican-based VS CNI-based therapy. Transplant International. 2011;14:22.

Song G, Chen CL, Lee WC, Joh JW, Levy G, Lopez P, et al., editors. Design of a randomized study aimed to evaluate the efficacy and safety of everolimus with reduced exposure tacrolimus versus standard tacrolimus in living donor liver transplant recipients. ILTS; 2013.

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.

University of York. Centre for Reviews and Dissemination, & Akers, J. (2008).Systematic reviews: CRD's guidance for undertaking reviews in health care. Centre for Reviews and Dissemination.

Wiesner RH, Menon KV. Late hepatic allograft dysfunction. Liver Transpl. 2001;7(11 Suppl 1):S60-73.