National Institute for Health and Care Excellence

Draft for consultation

Osteoporosis: risk assessment

[C] Validity of fragility fracture risk prediction tools

NICE guideline <number>

Evidence review underpinning recommendations 1.3.1-1.3.9 and recommendation for research in the NICE guideline

January 2026

Draft for Consultation

These evidence reviews were developed by NICE

1

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN:

Contents

1

2

1. Validity of fragility fracture risk prediction tools

1.1. Review questions: What is the validity of risk prediction tools for predicting the risk of fragility fractures in adults, including those who have had a previous fragility fracture?

1.1.1. Introduction

This review aims to look at the performance of UK-validated fragility fracture risk prediction tools regarding their overall fit, calibration, and discriminatory power in adults at risk or suspected risk of fragility fracture.

1.1.2. Summary of the protocol

For full details see the review protocols in Appendix A.

Table 1: PI(C)OTS characteristics of review question

Population	Adults (18 years and older) who are at suspected risk of fragility fractures (people with or at risk of primary or secondary osteoporosis or have had a previous fragility fracture).

Intervention	**Risk prediction tools**
	Risk of major osteoporotic fracture or hip fracture using:
	• CFracture
	• FRAX-UK/FRAXplus-UK
	  ○ Without bone mineral density (BMD) assessment
	  ○ With BMD
	  ○ With BMD and trabecular bone score (TBS)
	• FRAX with NOGG thresholds
	• IDFracture
	• QFracture
	Strata: Version or iteration or risk prediction tool; type of fracture.

Outcomes	**Risk prediction tools** • Overall fit: R^2 statistic, Brier score • Calibration: calibration plots and curves; calibration in the large; observed:expected ratio; integrated calibration index • Discrimination: c-statistic/AUC, D statistic for overall discrimination • Reclassification (for example, net reclassification index) • Discrimination at specific threshold: sensitivity/specificity, predictive values
Target condition	Fragility fracture • Major osteoporotic fracture • Hip fracture
Setting	• Any setting
Study design	• Internal or external prospective or retrospective cohort validation studies

Abbreviations: AUC, area under the curve; FRAX, Fracture Risk Assessment Tool; NOGG, National Osteoporosis Guideline Group.

1.1.3. Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual. Methods specific to this review question are described in the review protocol in appendix A and in Section 1.2 below.

Declarations of interest were recorded according to NICE's conflicts of interest policy.

1.1.4. Risk prediction evidence

Evidence was identified regarding risk prediction tools for predicting the risk of fragility fractures in adults including those who have had a previous fragility fracture. The assessments and the specific outcomes are summarised in Table 4. Full details can be found in Appendix D, Appendix E, and Appendix F.

Evidence on the accuracy of bone assessment methods to predict fragility fracture was sought as part of Evidence review D, whilst evidence on the effectiveness of risk assessment tools was sought as part of Evidence review E.

1.1.5. Fragility fracture risk prediction tools

For a list of the predictors (risk factors) included in the risk prediction tools and model features, see Table 2 and Table 3.

1.1.5.1. CFracture

CFracture was developed in 2023 and has been internally validated based on a large general primary care population in the UK (CPRD GOLD database). The algorithm uses the same risk factors, which are readily available in electronic healthcare records, as the QFracture tool. CFracture estimates 10-year risk of hip or major osteoporotic (hip, shoulder, spine, or wrist) fracture in men and women aged 30-99 years. Ascertainment of fracture (that is, whether a fracture has occurred) was through use of primary care Read codes from the CPRD GOLD database, and linked Hospital Episodes Statistics (HES) admitted patient care or Office for National Statistics (ONS) death registration (both of which use ICD-10 codes).

The clinical risk factors included in the tool are: age, sex, ethnicity, smoking status, alcohol status, type 1 and type 2 diabetes, body mass index (BMI), parental history of osteoporosis/hip fracture, resident in a nursing or care home, previous fragility fracture, history of falls, dementia, cancer, asthma, chronic obstructive pulmonary disease, cardiovascular disease, chronic kidney disease, epilepsy, Parkinson's disease, systemic lupus erythematosus, rheumatoid arthritis, chronic liver disease, gastrointestinal conditions likely to result in malabsorption (Crohn's disease, ulcerative colitis, celiac disease, steatorrhoea, blind loop syndrome), other endocrine conditions (thyrotoxicosis, primary or secondary hyperparathyroidism, Cushing's syndrome), ≥2 prescriptions for systemic corticosteroids in the six months prior to cohort entry, ≥2 prescriptions for antidepressants six months prior to cohort entry. In women, ≥2 prescriptions for oestrogen-only hormone replacement therapy six months prior to cohort entry is also included in the model.

One strength of the CFracture model is that it uses a Fine-Grey hazard model and associated Aalen-Johansen risk estimator to account for the competing risk of death from non-fracture causes. As such, the model does not assume that individuals lost to follow up have the same risk as those who are not.

1.1.5.2. FRAX/FRAXplus

The FRAX tool was developed in 2008 and was developed using baseline and follow up data from nine prospective population-based cohorts (including Europe, Australia, Canada and Japan) and validated in 11 prospective population-based cohorts (>1 million patient years),

all of which had similar risk profiles to the development cohorts. The tool estimates 10-year hip or major osteoporotic (hip, shoulder, spine, or wrists) fracture risk in women and men aged 40-90 years and can be used with or without a BMD measurement. For clarity, in this guideline we have used the terms 'FRAX with BMD' and 'FRAX without BMD.' Ascertainment of fracture was through self-report or hospital or central databases depending on the cohort.

The clinical risk factors included in the FRAX algorithm are: age, sex, weight, height, previous fracture, parental hip fracture, alcohol use, current smoking, glucocorticoids, and rheumatoid arthritis.

To be used in a specific country, the tool needs to be calibrated using country-specific fracture incidence and mortality data. The tool is used widely across the world with currently more than 80 country-specific models available on its website (www.fraxplus.org). A UK version of the FRAX tool was calibrated in 2006 using fracture incidence and mortality data from 1998.

One strength of the FRAX model, unlike QFracture (which uses a Cox model and Kaplan-Meier risk estimate), is that it uses a Poisson regression model and associated maximum likelihood estimator to account for the competing risk of death from non-fracture causes.

1.1.5.3. QFracture

QFracture was developed in 2009 and has been internally (Hippisley-Cox 2009) and externally validated (Collins 2011) in large UK general primary care populations (QResearch and THIN clinical databases). The algorithm is based on variables that are readily available in electronic healthcare records and provides an estimate of an individual's 10-year risk of first incident hip fracture or the 10-year risk of first incident major osteoporotic (hip, spine, and wrist) fractures (including without the need for a BMD measurement). It can be used in men or women aged 30−85 years without a previous fracture. Fracture ascertainment was through primary care Read codes from the QResearch database.

The clinical risk factors included in the QFracture algorithm in men and women are: age, sex, BMI, alcohol use, smoking, fall history, asthma, glucocorticoid use, cardiovascular disease, chronic liver disease, rheumatoid arthritis, type 2 diabetes, tricyclic antidepressant use. Additional factors used in women only are: hormone replacement therapy, parental history of hip fracture, menopausal symptoms, gastrointestinal malabsorption, and other endocrine disorders.

An updated version of the tool, QFracture 2012, has been internally (Hippisley-Cox 2012) and externally (Hippisley-Cox 2014, Livingstone 2022) validated in similarly large UK primary care population (QResearch and CPRD databases). This updated version included shoulder (i.e. proximal humerus) fractures in its definition of MOF, expanded the age range to 30-100 years, and added the following additional risk factors to the model: ethnicity, previous fracture, care home resident, use of antidepressants other than tricyclic antidepressants, use of anticonvulsants, cancer, chronic renal disease, COPD, dementia, epilepsy, Parkinson's disease, systemic lupus erythematosus, and Type 1 diabetes. The current version of QFracture is the 2016 version, which improves ascertainment of fracture by using linked Hospital Episode Statistics (HES) data. However, there has not yet been a published internal validation study and its performance in the general population is only briefly summarised on the QFracture website (www.qfracture.org).

One major drawback of the QFracture model is that it does not account for competing risks because it uses a Cox proportional hazards model and provides an associated Kaplan-Meier risk estimate. As such, because the model assumes that individuals who are lost to follow up have the same fracture risk as those who are not lost to follow up, it will systematically overpredict fracture risk.

Table 2: Risk factors (predictors) included in the fragility fracture risk prediction tools

Risk factor	CFracture	FRAX®/FRAXplus®	QFracture
Age (years) in development cohort	30-99	40-90	2009: 30-85 2012/2016: 30-100
Sex	Y	Y	Y
BMI	Y	Y	Y
Weight	Y	Y	Y
Height	Y	Y	Y
Bone mineral density **(femoral neck T-score or absolute value)**	-	Optional FRAXplus: optional, lumbar spine BMD included	-
Trabecular bone score	-	Derived model	-
Other anthropometric parameters	-	FRAXplus: Y Hip axis length	-
Ethnicity	Y	US and Singapore versions only	2012/2016
Fracture history	Y	Y FRAXplus: Y recency according to site	2012/2016
Parental history of osteoporosis or hip fracture	Y	Y	Y
Fall history	Y	FRAXplus	Y
Resident in nursing or care home	Y	-	2012/2016
Secondary osteoporosis[a] (Yes/No)	-	Y	-
Smoking	Y	Y	Y
Alcohol use	Y	Y	Y
Antidepressant use	Y	TCA only	2009: TCA only 2012/2016: use of antidepressants other than TCA included as separate variable

Risk factor	CFracture	FRAX®/FRAXplus®	QFracture
Glucocorticoid use	Y	Y FRAXplus: Y high dose use	Y
Hormone replacement therapy	Y, oestrogen-only	-	2012/2016
Asthma	Y	-	Y
Cancer	Y	-	2012/2016
Cardiovascular disease	Y	-	Y
Chronic kidney disease	Y	-	2012/2016
Chronic liver disease	Y	-	Y
COPD	Y	-	2012/2016
Dementia	Y	-	2012/2016
Diabetes	Y (T1DM and T2DM)	FRAXplus: Y duration of T2DM	Y
Endocrine disorders	Y	-	Y
Epilepsy or anticonvulsant use	Y	-	2012/2016
Gastrointestinal malabsorption[b]	Y	-	Y
Menopausal symptoms	Y	-	Y
Parkinson's disease	Y	-	2012/2016
Rheumatoid arthritis	Y	Y	Y
Systematic lupus erythematosus	Y	-	2012/2016

Abbreviations: COPD, chronic pulmonary obstructive disease; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus; TCA, tricyclic antidepressants.

Notes

a. For example: Type 1 diabetes, chronic hyperthyroidism, premature menopause, chronic liver disease, chronic malnutrition, chronic liver disease.

b. Includes Crohn's disease, ulcerative colitis, celiac disease, steatorrhoea, blind loop syndrome.

c. Including taking progesterone.

d. For example: Down's syndrome, other syndromal disorders, visual impairment, treatment with antipsychotics.

Table 3: Model features of fragility fracture risk prediction tools

Fracture risk prediction tool	Fracture ascertainment	Type of statistical model: associated risk estimator	Accounts for competing risks?
CFracture	CPRD GOLD Read codes + linked HES + ONS data	Fine-Grey sub-distribution hazard	Yes

Fracture risk prediction tool	Fracture ascertainment	Type of statistical model: associated risk estimator	Accounts for competing risks?
		model: Aalen-Johansen	
FRAX	Self-report or hospital/central databases	Poisson regression generalised linear model; Maximum likelihood estimation	Yes
QFracture	QResearch: QResearch Read codes 2016 version: QResearch Read codes + linked HES data	Cox's proportional hazards model; Kaplan-Meier	No

Abbreviations: CPRD, Clinical Practice Research Database; HES, Hospital Episode Statistics; ONS, Office of National Statistics.

1.1.6. Included studies

Risk prediction tool review

Ten validation studies on three fragility fracture risk prediction tools were included in the review. Nine of the studies were prospective cohort studies, whilst one study (Green 2024) was a retrospective cohort study. Seven studies were external validation studies: Akyea 2019 (FRAX-UK without BMD and QFracture 2016), Collins 2011 (QFracture 2009), Green 2024 (FRAX-UK with and without BMD), Hippisley-Cox 2014 (QFracture 2012), Ihama 2021 (FRAX-UK without BMD and QFracture 2016), Klop 2016 (FRAX-UK without BMD), and Livingstone 2022 (QFracture 2012). One study (Hippisley-Cox 2012) reported an internal validation study only of QFracture 2012. One study reported an internal validation study of one tool and an external validation study of another tool in the same population cohorts (Hippisley-Cox 2009: QFracture 2009 and FRAX-UK respectively) allowing direct comparison of the tools.

Five of the external validation studies were conducted in the general population and used large UK general primary care population databases of people registered with a general practitioner (Akyea 2019, Collins 2011, Hippisley-Cox 2013, Klop 2016, Livingstone 2022). Two studies were conducted in settings other than primary care: Ihama 2021 was conducted in 18 care homes in Lincolnshire, UK, whilst Green 2024 was conducted in a tertiary hospital in Sheffield, UK. Three studies were conducted in people with comorbidities associated with an increased risk of fragility fracture, including people with: chromic pulmonary obstructive disease (Akyea 2019), coeliac disease (Green 2024), and rheumatoid arthritis (Klop 2016).

The included studies for the risk prediction tool review are summarised in Table 4 below. Evidence from these studies is summarised in the clinical evidence summary below. Meta-analysis of the risk prediction tools was not conducted due to the small number of studies for each version of the risk prediction tools and the different populations they were conducted in. Published calibration plots from the included studies were presented to the guideline committee for consideration and if not reported, overall observed: expected (O:E) ratios were estimated from them.

See also the study selection flow chart in Appendix C, study evidence tables in Appendix D, forest plots in Appendix E, and PROBAST risk of bias tables in Appendix F.

1.1.7. Excluded studies

See the excluded studies list in Appendix G.

1.1.8. Summary of studies included in the fragility fracture risk prediction tool evidence review

Table 4: Summary of studies included in the fragility fracture risk prediction tool evidence review

Study **Type of study**	**Total number of participants (Men/Women)** **Age in years** **Length of FU in years**	**Validation cohort** **Setting, population**	**Risk prediction tool**	**Outcomes predicted**	**Overall fit, calibration, and discrimination measures**	**Overall risk of bias** **Overall directness**
Akyea 2019 External prospective cohort	N=80,874 (42,799/38,075) eligible participants at baseline N=72,559 validation dataset Mean age: 66.9 (SD 10) Median length of FU: COPD patients 5.28 (IQR 2.6-8.3); Non-COPD patients 5.24 (IQR 2.6-8.3)	THIN database Primary care, GP-registered people with COPD≥40 years-old	- FRAX-UK, v.3.12 - QFracture 2016	- 10-year MOF - 10-year HF	**Discrimination** - AUC Threshold at ≥20% (MOF) and ≥3% (HF) risk: - Sensitivity - Specificity - PPV - NPV	- High risk of bias - Directly applicable
Collins 2011 External prospective cohort	N=2,244,636 (1,108,219/1,136,417)	THIN database Primary care, GP-registered, 30-85 years-old	- QFracture 2009	- 10-year OF - 10-year HF	**Overall fit** - R^2 - Brier score **Calibration** - Calibration plot	- High risk of bias - Partially applicable

Study Type of study	Total number of participants (Men/Women) Age in years Length of FU in years	Validation cohort Setting, population	Risk prediction tool	Outcomes predicted	Overall fit, calibration, and discrimination measures	Overall risk of bias Overall directness
	Median age: Men 47 (IQR 37-59); Women 48 (37-62) Median length of FU: OF: 5.98 (IQR 2.61-8.5). HF: 6.03 (IQR 2.62-8.5)				**Discrimination** - AUC - D-statistic	
Green 2024 External retrospective cohort	N=593 (187/406) Median age: 45.0 (IQR 31.5-57.6) Median length of FU: 10.5 (IQR 9.0-13.4)	Coeliac disease + DXA scan databases Tertiary hospital, people with biopsy-proven coeliac disease	- FRAX-UK with BMD - FRAX-UK without BMD	- 10-year MOF	**Discrimination** NOGG age-specific thresholds: - Sensitivity - Specificity - PPV - NPV	- High risk of bias - Directly applicable
Hippisley-Cox 2009 Internal prospective cohort for QFracture 2009 External prospective cohort for FRAX-UK	N=1,275,917 (633,764/642,153) Median age: Men: 46.0 (IQR 37-69); Women: 49.0 (IQR 37-63) Median length of FU: MOF: Men 5.6 (2.2-10.4); Women 5.7 (2.2-10.5). HF:	QResearch (version 20) database Primary care, GP-registered, 30-85 years-old	- QFracture 2009 - FRAX-UK	- 10-year risk OF - 10-year HF	**Overall fit** - R^2 **Calibration** - Calibration plot - O-E ratio **Discrimination** - AUC - D-statistic	- High risk of bias - Partially applicable

Study Type of study	Total number of participants (Men/Women) Age in years Length of FU in years	Validation cohort Setting, population	Risk prediction tool	Outcomes predicted	Overall fit, calibration, and discrimination measures	Overall risk of bias Overall directness
	Men 5.7 (2.2-10.4); Women 5.9 (2.2-10.6).					
Hippisley-Cox 2012 Internal prospective cohort for QFracture 2012 External prospective cohort for QFracture 2009	N=1,583,373 (778,810/804,563) Age range: 30-100 Length of FU: 7.4[d]	QResearch (version 32) database Primary care, GP-registered, 30-100 years-old	- QFracture 2012	- 10-year MOF - 10-year HF	**Overall fit** - R^2 **Calibration** - Calibration plot - Observed, expected at ≥90% risk **Discrimination** - AUC - D-statistic Threshold at top decile of risk: - Sensitivity	- High risk of bias - Partially applicable
Hippisley-Cox 2014 External prospective cohort	N=3,271,512 (1,588,803/1,682,709) N=2,852,381 QFracture eligible patients in CPRD database Age range: 30-99	CPRD database Primary care, GP-registered, 30-99 years-old	- QFracture 2012	- 10-year MOF - 10-year HF	**Overall fit** - R^2 **Discrimination** - AUC - D-statistic Threshold at top decile of risk: - Sensitivity - Specificity - PPV	- High risk of bias - Partially applicable

Study Type of study	Total number of participants (Men/Women) Age in years Length of FU in years	Validation cohort Setting, population	Risk prediction tool	Outcomes predicted	Overall fit, calibration, and discrimination measures	Overall risk of bias Overall directness
	Length of FU: Up to 14.5				- NPV	
Ihama 2021 External prospective cohort	N= 217 (83/124) Mean age: 81.21 (SD 12.51) Length of FU: 12 months	Care homes in Boston, Lincolnshire, UK Adult care home residents	- FRAX-UK without BMD - QFracture 2016	- 10-year MOF	**Discrimination** - c-statistic	- High risk of bias - Directly applicable
Klop 2016 External prospective cohort	<u>Recalibration</u> N=11,582 (3729/7853) people with rheumatoid arthritis N=38,755 people (matched cohort from general population); N=24,227 people (matched cohort from general population after HES linkage)	<u>Recalibration</u> CPRD GOLD database Primary care, GP registered, 40-90 years-old with RA Primary care, GP-registered, 40-90 years-old (matched cohort) <u>Extension</u>	- FRAX-UK (v.3.9) without BMD	- 10-year MOF - 10-year HF	**Calibration** - Calibration plots **Discrimination** - c-statistic **Reclassification** - Net reclassification index	- High risk of bias - Directly applicable (RA population)/Partially applicable (matched cohort)

Study **Type of study**	**Total number of participants (Men/Women)** **Age in years** **Length of FU in years**	**Validation cohort** **Setting, population**	**Risk prediction tool**	**Outcomes predicted**	**Overall fit, calibration, and discrimination measures**	**Overall risk of bias** **Overall directness**
	Mean age: 62.9 (SD 11.4) <u>Extension</u> N=7,221 people with rheumatoid arthritis (2263/4958) Age: NR Median length of FU (IQR): 9.0 (5.3-10.0)	CPRD GOLD database with linked HES data Primary care, GP registered, 40-90 years-old with RA				
Livingstone 2022 External prospective cohort	N=5,432,139 (2,684,730/2,747,409) Mean age (SD): Men 48.5 (15.6); Women 50.7 (17.4) Length of FU: Up to 12.2	CPRD GOLD database + linked HES and ONS mortality data Primary care, GP-registered, 30-100 years-old	- QFracture 2012 - QFracture 2012 extension with competing risks	- 10-year MOF - 10-year HF	**Overall fit** - R^2 **Calibration** - Calibration plot - O:E ratio **Discrimination** - c-statistic - D-statistic	- High risk of bias - Partially applicable
Livingstone 2023[e] Internal prospective cohort for CFracture	N=1,810,713 (894,910/915,803)	CPRD GOLD database + linked HES and ONS mortality data	- CFracture	- 10-year MOF - 10-year HF	**Calibration** - Calibration plot - O:E ratio **Discrimination**	- High risk of bias - Partially applicable

Study Type of study	Total number of participants (Men/Women) Age in years Length of FU in years	Validation cohort Setting, population	Risk prediction tool	Outcomes predicted	Overall fit, calibration, and discrimination measures	Overall risk of bias Overall directness
	Median age: Men: 45 (IQR 35-59); Women: 47 (IQR 35-63) Length of FU: Up to 22.3	Primary care, GP-registered, 30-99 years-old			- c-statistic	

Abbreviations: AUC, area under the curve; COPD, chronic obstructive pulmonary disease; CPRD, Clinical Practice Research Datalink; EHR, electronic health record; FU, follow up; HES, Hospital Episode Statistics; HF, hip fracture; IQR, interquartile range MOF, major osteoporotic fracture (hip, vertebral, wrist, proximal humeral or osteoporotic fractures); NOGG, National Osteoporosis Guideline Group; NPV, negative predictive value; NR, not reported; OF, osteoporotic fracture (hip, vertebral, or distal radius fracture); ONS, Office of National Statistics; O:E, ratio of observed risk to mean predicted risk; PPV, positive predictive value; QCT, quantitative computed tomography; RA, rheumatoid arthritis; THIN, The Health Improvement Network.

Notes:

a. In addition to HES, CPRD Aurum is linked to Death Registration, Cancer data, Mental Health Services Dataset, and Small Area-level data.

b. Calculated from reported person-years of observation and number of people in cohort.

c. Size of internal validation cohort is unclear as insufficient detail provided about methods.

d. Fragility fracture risk prediction tool evidence: Overall fit, calibration, and discrimination data.

e. This study uses a 2:1 split of the same population cohort of Livingstone 2022 to develop and internally validate CFracture.

1.1.8.1. CFracture

Table 5: Clinical evidence profile: Performance data for 10-year risk of major osteoporotic fracture using CFracture

Study	Type of validation study Cohort, population	Sex	Overall fit	Calibration plot O:E ratio (95% CI)	Discrimination
Livingstone 2023	Internal prospective cohort CPRD GOLD primary care database, adults 30-99 years	Men	- NR	- Yes - 1.06 (0.98-1.15)	- c-statistic=0.738 (0.732-0.743)
Livingstone 2023	Internal prospective cohort CPRD GOLD primary care database, adults 30-99 years	Women	- NR	- Yes - 1.16 (1.11-1.21)	- c-statistic=0.813 (0.810-0.816)

Abbreviations: CPRD, Clinical Practice Research Datalink; O:E ratio, observed: expected ratio.

Table 6: Clinical evidence summary: Performance data for 10-year risk of hip fracture using CFracture

Study	Type of validation study Cohort, population	Sex	Overall fit	Calibration plot O:E ratio (95% CI)	Discrimination
Livingstone 2023	Internal prospective cohort CPRD GOLD, adults 30-99 years	Men	- NR	- Yes - 1.18 (1.05-1.32)	- c-statistic=0.886 (0.877-0.895)

Livingstone 2023	Internal prospective cohort CPRD GOLD primary care database, adults 30-99 years	Women	- NR	- Yes - 1.07 (0.98-1.15)	- c-statistic=0.914 (0.908-0.919)

Abbreviations: CPRD, Clinical Practice Research Datalink; O:E ratio, observed: expected ratio.

1.1.8.2. FRAX-UK/FRAXplus-UK

Table 7: Clinical evidence summary: Performance data for 10-year risk of major osteoporotic fracture using FRAX-UK

Study	Type of validation study Cohort, population	Sex	Overall fit	Calibration plot O:E ratio (95% CI)	Discrimination (95% CI)
Akyea 2019	External prospective cohort THIN primary care database, adults with COPD≥40 years-old	All	- NR	- NR	- AUC=0.714 (0.706-0.722) **Threshold ≥20% risk** - Sensitivity=25.4% (22.7-28.1) - Specificity=92.6% (91.0-94.2) - PPV=18.8% (16.4-21.1) - NPV=94.8% (93.4-96.2)

Study	Type of validation study Cohort, population	Sex	Overall fit	Calibration plot O:E ratio (95% CI)	Discrimination (95% CI)
Green 2024	External retrospective cohort Single-centre, tertiary hospital, Sheffield, UK, adults≥18 years-old with biopsy-proven coeliac disease	All	- NR	- NR	**Without BMD using NOGG age-specific thresholds:**[b] - Sensitivity=22.0 % (12.0-35.0) - Specificity=91.0% (89.0-94.0) - PPV=16.3% (8.7-27.6) - NPV=93.5% (91.1-95.3) **With BMD using NOGG age-specific thresholds:**[b] - Sensitivity=15% (6.0-27.0) - Specificity=92.0% (89.0-94.0) - PPV=11.4% (4.9-22.6) - NPV=93.4% (91.0-95.2)
Ihama 2021	External prospective cohort Care homes in Lincolnshire, UK, adult care home residents	All	- NR	- NR	- c-statistic=0.655 (0.469-0.803)[c]
Klop 2016	External prospective cohort CPRD primary care database, adults 40-90 years	All	- NR	- Yes - NR	- c-statistic=0.71 (0.698-0.722)[c]

Study	Type of validation study Cohort, population	Sex	Overall fit	Calibration plot O:E ratio (95% CI)	Discrimination (95% CI)
Klop 2016	External prospective cohort CPRD primary care database, adults with RA, 40-90 years	All	- NR	- Yes - 0.632 (0.558-0.706)[a]	- c-statistic=0.69 (0.671-0.708)[c]

Abbreviations: COPD, chronic obstructive pulmonary disease; CPRD, Clinical Practice Research Datalink; NOGG, National Osteoporosis Guideline Group; O:E ratio, observed: expected ratio; RA, rheumatoid arthritis.
Notes:
a. O:E ratio and 95% CIs calculated using equations in Table 2, Debray 2018.
b. 95% CIs calculated by developers.
c, 95% CI calculated from reported total number of patients, number of incident fractures, and number of patients not experiencing incident fractures using equations in Table 2, Debray 2018.

Table 8: Clinical evidence summary: Performance data for 10-year risk of hip fracture using FRAX-UK

Study	Type of validation study Cohort, population	Sex	Overall fit	Calibration plot O:E ratio (95% CI)	Discrimination
Akyea 2019	External prospective cohort THIN primary care database, adults with COPD≥40 years	All	- NR	- NR	- AUC=0.761 (0.749-0.772) **Threshold≥3% risk** - Sensitivity=78.1% (75.6-80.7) - Specificity=60.8% (57.8-63.8) - PPV=3.9% (2.7-5.1) - NPV=99.3% (98.8-99.8)

Hippisley-Cox 2009	External prospective cohort QResearch primary care database, adults 40-85 years	Men	- R^2=54.07% (52.1-53.65)	- Yes - 0.741 (0.673-0.808)[a]	- AUC=0.817 (0.807-0.827)[c] - D-statistic=2.22 (2.14-2.3)
Hippisley-Cox 2009	External prospective cohort QResearch primary care database, adults 40-85 years	Women	- R^2=54.83% (54.43-55.12)	- Yes - 0.868 (0.815-0.921)[a]	- AUC=0.845 (0.840-0.850)[c] - D-statistic=2.26 (2.21-2.3)
Klop 2016	External prospective cohort CPRD primary care database, adults 40-90 years	All	- NR	- Yes - 0.884 (0.773-0.995)[a]	- c-statistic=0.83 (0.812-0.847)[c]
Klop 2016	External prospective cohort CPRD primary care database, adults with RA, 40-90 years	All	- NR	- Yes <u>Calibration</u> - 0.748 (0.561-0.935)[a] <u>Recalibration</u> - 0.748 (0.511-0.985)[a] <u>Extension</u> - 0.943 (0.649-1.238)[a]	- c-statistic=0.78 (0.752-0.805)[c]

Abbreviations: COPD, chronic obstructive pulmonary disease; CPRD, Clinical Practice Research Datalink; NOGG, National Osteoporosis Guideline Group; O:E ratio, observed: expected ratio; RA, rheumatoid arthritis.

Notes:

a. O:E ratio calculated using equations in Table 2, Debray 2018.b. 95% CI not estimable as only 9 datapoints available on the published calibration plot.

c. 95% CI calculated from reported total number of patients, number of incident fractures, and number of patients not experiencing incident fractures using equations in Table 2, Debray 2018.

1.1.8.2.1. Reclassification: FRAX-UK recalibrated versus extended model

Table 9: Reclassification by FRAX-UK extended model for adults with rheumatoid arthritis

Study	Comparison	Description	Reclassification
Klop 2016	Extended FRAX-UK compared to FRAX-UK recalibrated using NOGG intervention age-specific thresholds	Extended model includes duration of rheumatoid arthritis, high-dose glucocorticoid use, and secondary osteoporosis as predictors, in addition to those included in FRAX-UK.	NRI=0.01 (95% CI -0.04-0.05)

1.1.8.3. QFracture-2009

Table 10: Clinical evidence summary: Performance data for 10-year risk of major osteoporotic fracture using QFracture 2009

Study	Type of validation study Cohort, population	Sex	Overall fit (95% CI)	Calibration plot O:E ratio (95% CI)	Discrimination (95% CI)
Hippisley-Cox 2009	Internal prospective cohort QResearch (version 20) primary care database, adults 30-85 years	Men	- R^2=30.02% (22.21-37.84)	- Yes - 0.984 (0.953-1.014)[a]	- AUC=0.688 (0.684-0.692) - D-statistic=1.34 (1.09-1.59)
Hippisley-Cox 2009	Internal prospective cohort QResearch (version 20) primary care database, adults 30-85 years	Women	- R^2=44.87% (43.07-46.67)	- Yes - 0.999 (0.975-1.023)[a]	- AUC=0.788 (0.786-0.790) - D-statistic=1.85 (1.78-1.91)
Collins 2011	External prospective cohort THIN primary care database, adults 30-85 years	Men	- R^2=37.99% (36.64-39.35) - Brier=0.010 (0.008-0.012)	- Yes - 0.953 (0.929-0.978)[a]	- AUC=0.739 (0.733-0.745)[b] - D-statistic=1.60 (1.56-1.65)
Collins 2011	External prospective cohort THIN primary care database, adults 30-85 years	Women	- R^2=49.24% (48.64-49.85) - Brier=0.027 (0.025-0.029)	- Yes - 0.950 (0.929-0.971)[a]	- AUC=0.816 (0.813-0.819)[b] - D-statistic=2.02 (1.99-2.04)

Abbreviations: THIN, The Health Improvement Network; O:E ratio, observed: expected ratio.

Notes:
a. O:E ratio calculated using equations in Table 2, Debray 2018.
b. 95% CI calculated from reported total number of patients, number of incident fractures, and number of patients not experiencing incident fractures using equations in Table 2, Debray 2018.

Table 11: Clinical evidence summary: Performance data for 10-year risk of hip fracture using QFracture 2009

Study	Type of validation study Cohort, population	Sex	Overall fit	Calibration plot O:E ratio (95% CI)	Discrimination
Hippisley-Cox 2009	Internal prospective cohort QResearch (version 20) primary care database, adults 30-85 years QResearch (version 20) primary care database, adults 40-85 years	Men	- R^2=63.19% (60.81-65.57)	- Yes Adults 30-85 years - 0.879 (0.767-0.992)[a] Adults 40-85 years - 0.906 (0.817-0.994) [a]	- AUC=0.856 (0.851-0.860) - D-statistic=2.68 (2.55-2.82)
Hippisley-Cox 2009	Internal prospective cohort QResearch (version 20) primary care database, adults 30-85 years QResearch (version 20) primary care database, adults 40-85 years	Women	- R^2=63.94% (62.12-65.76)	- Yes Adults 30-85 years - 0.968 (0.852-1.084)[a] Adults 40-85 years - 0.982 (0.900-1.065)[a]	- AUC=0.890 (0.889-0.892) - D-statistic=2.73 (2.62-2.83)
Collins 2011	External prospective cohort THIN primary care database, adults 30-85 years	Men	- R^2=60.42% (59.22-61.63) - Brier=0.005 (0.003-0.007)	- Yes - OE ratio not extractable[b]	- AUC=0.855 (0.848-0.862)[c] - D-statistic=2.53 (2.46-2.59)
Collins 2011	External prospective cohort THIN primary care database, adults 30-85 years	Women	- R^2=62.82% (62.22-63.43) - Brier=0.013 (0.012-0.015)	- Yes - OE ratio not extractable[b]	- AUC=0.890 (0.887-0.893)[c] - D-statistic=2.66 (2.63-2.70)

Abbreviations: THIN, The Health Improvement Network; O:E ratio, observed: expected ratio.

Notes:

a. O:E ratio calculated using equations in Table 2, Debray 2018.

b. Data for observed and estimated risk not extractable from calibration plot.

c. 95% CI calculated from reported total number of patients, number of incident fractures, and number of patients not experiencing incident fractures using equations in Table 2, Debray 2018.

1.1.8.4. QFracture-2012

Table 12: Clinical evidence summary: Performance data for 10-year risk of major osteoporotic fracture using QFracture-2012

Study	Type of validation study Cohort, population	Sex	Overall fit	Calibration plot O:E ratio (95% CI)	Discrimination
Hippisley-Cox 2012	Internal prospective cohort QResearch (version 32) primary care database, adults 30-100 years	Men	- R^2=38.20% (36.89-39.57)	- Yes - 0.866 (0.841-0.891)[a]	- AUC=0.711 (0.703-0.719) - D-statistic=1.61 (1.56-1.66) **Threshold≥90% risk:** - Sensitivity=37.0% (36.0-38.0) - Specificity=not estimable
Hippisley-Cox 2012	Internal prospective cohort QResearch (version 32) primary care database, adults 30-100 years	Women	- R^2=51.9% (51.2-52.6)	- Yes - 0.897 (0.876-0.917)[a]	- AUC=0.790 (0.787-0.793) - D-statistic=2.13 (2.10-2.15) **Threshold≥90% risk**: - Sensitivity=35.0% (34.0-36.0) - Specificity=not estimable
Hippisley-Cox 2014	External prospective cohort CPRD primary care database, adults 30-99 years	All	- NR	- NR	**Threshold≥90% risk:** - Sensitivity=50.0% (49.0-50.0) - Specificity=90.0% (90.0-91.0)
Hippisley-Cox 2014	External prospective cohort	Men	- R^2=49.8% (48.9-50.7)	- Yes - 0.744 (0.722-0.766)[a]	- AUC=0.768 (0.763-0.773) - D-statistic=2.038 (2.002-2.075)

Study	Type of validation study Cohort, population	Sex	Overall fit	Calibration plot O:E ratio (95% CI)	Discrimination
	CPRD primary care database, adults 30-99 years				
Hippisley-Cox 2014	External prospective cohort CPRD primary care database, adults 30-99 years	Women	- R^2=56.3% (55.8-56.7)	- Yes - 0.823 (0.807-0.839) [a]	- AUC=0.817 (0.814-0.819) - D-statistic=2.322 (2.301-2.343)
Livingstone 2022	External prospective cohort CPRD GOLD primary care database, adults 30-100 years	Men	- R^2=42.4% (41.9-43.0)	- Yes - 1.817 (1.806-1.827) without competing risks[a] - 1.483 (1.473-1.494) with competing risks[a]	- c-statistic=0.738 (0.735-0.741) - D-statistic=1.76 (1.74-1.78)
Livingstone 2022	External prospective cohort CPRD GOLD primary care database, adults 30-100 years	Women	- R^2=54.8% (54.5-55.1)	- Yes - 1.508 (1.481-1.536) without competing risks[a] - 1.212 (1.185-1.239) with competing risks[a]	- c-statistic=0.813 (0.811-0.815) - D-statistic=2.25 (2.24-2.27)

Abbreviations: CPRD, Clinical Practice Research Datalink; THIN, The Health Improvement Network; O:E ratio, observed: expected ratio.
Notes:
a. O:E ratio and 95% CIs calculated using equations in Table 2, Debray 2018.

Table 13: Clinical evidence summary: Performance data for 10-year risk of hip fracture using QFracture-2012

Study	Type of validation study Cohort, population	Sex	Overall fit	Calibration plot O:E ratio (95% CI)	Discrimination
Hippisley-Cox 2012	Internal prospective cohort QResearch (version 32) primary care database, adults 30-85 years	Men	- R^2=70.37% (69.25-71.49)	- Yes - 0.785 (0.732-0.839)[a]	- AUC=0.875 (0.868-0.883) - D-statistic=3.15 (3.06-3.24) **Threshold≥90% risk:** - Sensitivity=64.0% (62.0-67.0) - Specificity=not estimable
Hippisley-Cox 2012	Internal prospective cohort QResearch (version 32) primary care database, adults 30-85 years	Women	- R^2=71.73% (71.0-72.30)	- Yes - 0.799 (0.749-0.850)[a]	- AUC=0.893 (0.890-0.896) - D-statistic=3.26 (3.21-3.31) **Threshold≥90% risk:** - Sensitivity=60.0% (58.0-61.0) - Specificity=not estimable
Hippisley-Cox 2014	External prospective cohort CPRD primary care database, adults 30-99 years	All	- NR	- NR	**Threshold≥90% risk:** - Sensitivity=67.0% (66.0-67.0) - Specificity=90.0% (90.0-91.0)
Hippisley-Cox 2014	External prospective cohort CPRD primary care database, adults 30-99 years	Men	- R^2=69.0% (68.5-70.0)	- Yes - 0.765 (0.712-0.817)[a]	- AUC=0.872 (0.867-0.877) - D-statistic=2.046 (1.977-2.116)

Hippisley-Cox 2014	External prospective cohort CPRD primary care database, adults 30-99 years	Women	- R^2=70.6% (70.2-71.0)	- Yes - 0.859 (0.805-0.912)[a]	- AUC=0.890 (0.888-0.892) - D-statistic=3.171 (3.139-3.203)
Livingstone 2022	External prospective cohort CPRD GOLD primary care database, adults 30-100 years	Men	- R^2= 70.9% (70.4-71.3)	- Yes - 1.757 (1.720-1.793) with no competing risks[a] - 1.319 (1.288-1.349) with competing risks[a]	- c-statistic=0.888 (0.882-0.893) - D-statistic=3.19 (3.16-3.23)
Livingstone 2022	External prospective cohort CPRD GOLD primary care database, adults 30-100 years	Women	- R^2=71.7% (71.4-71.9)	- Yes - 1.306 without competing risks[a,b] - 0.930 with competing risks[a,b]	- c-statistic=0.918 (0.915-0.921) - D-statistic=3.26 (3.24-3.28)

Abbreviations: CPRD, Clinical Practice Research Datalink.
Notes:
a. O:E ratio and 95% CIs calculated using equations in Table 2, Debray 2018.
b. 95% CIs not estimable.

1.1.8.5. QFracture 2016

Table 14: Clinical evidence summary: Performance data for 10-year risk of major osteoporotic fracture using QFracture-2016

Study	Type of validation study Cohort, population	Sex	Overall fit	Calibration plot O:E ratio (95% CI)	Discrimination

Akyea 2019	External prospective cohort THIN primary care database, adults with COPD≥40 years-old	All	- NR	- NR	- AUC=0.614 (0.605-0.623) **Threshold≥20% risk:** - Sensitivity: 25.2% (22.5-27.9) - Specificity: 87.7% (85.7-89.7) - PPV: 12.2% (10.2-14.2) - NPV: 94.5% (93.1-95.9)
Ihama 2021	External prospective cohort Care homes, Bedfordshire, UK, adult care home residents	All	- NR	- NR	- c-statistic=0.736 (0.553-0.862)

Abbreviations: THIN, The Health Improvement Network; O:E ratio, observed: expected ratio.

Table 15: Clinical evidence summary: Performance data for 10-year risk of hip fracture using QFracture-2016

Study	Type of validation study Cohort, population	Sex	Overall fit	Calibration plot O:E ratio (95% CI)	Discrimination
Akyea 2019	External prospective cohort THIN primary care database, adults with COPD≥40 years-old	All	- NR	- NR	- AUC=0.761 (0.749-0.772) **Threshold≥3% risk:** - Sensitivity=82.1% (79.7-84.5) - Specificity=55.2% (52.1-58.3) - PPV=3.6% (2.5-4.8) - NPV=99.3% (98.8-99.8)

Abbreviations: COPD, chronic obstructive pulmonary disease; THIN, The Health Improvement Network; O:E ratio, observed:expected ratio.

1.1.9. GRADE profiles for discriminatory power of fragility fracture risk prediction tools

Table 16: GRADE profile for discriminatory power of risk prediction tools to predict 10-year risk of major osteoporotic fracture

Tool Population Type of validation study Validation cohort, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	c-statistic/AUC (95% CI)	GRADE certainty
CFracture Men 30-99 Internal prospective cohort - CPRD GOLD, HES, ONS (Livingstone 2023[a]), N=894,910	1	Very serious[b]	Not serious	Serious[c]	Not serious	- 0.738 (0.732-0.743)	VERY LOW
CFracture Women 30-99 Internal prospective cohort - CPRD GOLD, HES, ONS (Livingstone 2023 [a]), N=915,803	1	Very serious[b]	Not serious	Serious[c]	Not serious	- 0.813 (0.810-0.816)	VERY LOW

Tool Population Type of validation study Validation cohort, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	c-statistic/AUC (95% CI)	GRADE certainty
FRAX-UK without BMD Adults in care homes External prospective cohort UK regional care homes (Ihama 2021), N=207	1	Very serious[b]	Not serious	Not serious	Very serious[d]	- 0.655 (0.469-0.803)	VERY LOW
FRAX-UK without BMD Adults with COPD≥40 years-old External prospective cohort THIN (Akyea 2019), N=72,559	1	Very serious[b]	Not serious	Not serious	Not serious	- 0.714 (0.706-0.722)	LOW
FRAX-UK without BMD Adults 40-90 years External prospective cohort - CPRD primary care database with HES	1	Very serious[b]	Not serious	Not serious	Serious[d]	- 0.71 (0.698-0.722)	VERY LOW

Tool Population Type of validation study Validation cohort, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	c-statistic/AUC (95% CI)	GRADE certainty
linkage (Klop 2016), N=24,227							
FRAX-UK without BMD Adults with rheumatoid arthritis, 40-90 years <u>External prospective cohort</u> - CPRD primary care database (Klop 2016), N=7,221	1	Very serious[b]	Not serious	Not serious	Serious[d]	- 0.69 (0.671-0.708)	VERY LOW
QFracture 2009 Men 30-85 <u>Internal prospective cohort</u> QResearch v.20 (Hippisley-Cox 2009), N=633,764 <u>External prospective cohort</u> - THIN (Collins 2011), N=1,108,219	2	Very serious[b]	Very serious[e]	Serious[c]	Not serious	<u>Internal validation</u> - 0.688 (0.684-0.692) <u>External validation</u> - 0.739 (0.733-0.745)	VERY LOW

Tool Population Type of validation study Validation cohort, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	c-statistic/AUC (95% CI)	GRADE certainty
QFracture 2009 Women 30-85 <u>Internal prospective cohort</u> - QResearch v.20 (Hippisley-Cox 2009), N=642,153 <u>External prospective cohort</u> - THIN (Collins 2011), N=1,136,417	2	Very serious[b]	Serious[f]	Serious[c]	Not serious	<u>Internal validation</u> - 0.788 (0.786-0.790) <u>External validation</u> - 0.816 (0.813-0.819)	VERY LOW
QFracture 2012 Men 30-100 <u>Internal prospective cohort</u> - QResearch v.32 (Hippisley-Cox 2012), N=778,810	3	Very serious[b]	Serious[f]	Serious[c]	Not serious	<u>Internal validation</u> - 0.711 (0.703-0.719) <u>External validation</u> - 0.768 (0.763-0.773) HC2014 - 0.738 (0.735-0.741) LG2022	VERY LOW

Tool Population Type of validation study Validation cohort, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	c-statistic/AUC (95% CI)	GRADE certainty
<u>External prospective cohorts</u> - CPRD (Hippisley-Cox 2014), N=1,588,803 - CPRD GOLD, HES, ONS (Livingston 2022), N=2,684,730							
QFracture 2012 Women 30-100 <u>Internal prospective cohort</u> - QResearch v.32 internal validation cohort (Hippisley-Cox 2012), N=804,563 <u>External prospective cohorts</u> - CPRD (Hippisley-Cox 2014), N=1,682,709 - CPRD GOLD, HES, ONS (Livingstone 2022), N=2,747,409	3	Very serious[b]	Serious[f]	Serious[c]	Not serious	<u>Internal validation</u> - 0.790 (0.787-0.793) <u>External validation</u> - 0.817 (0.814-0.819) HC 2014 - 0.813 (0.811-0.815) LG2022	VERY LOW

Tool Population Type of validation study Validation cohort, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	c-statistic/AUC (95% CI)	GRADE certainty
QFracture 2016 Adults with COPD≥40 years-old <u>External prospective cohort</u> - THIN (Akyea 2019), N=72,559	1	Very serious[b]	Not serious	Not serious	Not serious	- 0.614 (0.605-0.623)	LOW
QFracture 2016 Adults in care homes <u>External prospective cohort</u> - Regional UK care homes external validation cohort (Ihama 2021), N=207	1	Very serious[b]	Not serious	Not serious	Serious[d]	- 0.736 (0.553-0.862)	VERY LOW

Abbreviations: COPD, chronic obstructive pulmonary disease; CPRD, Clinical Practice Research Datalink; HES, Hospital Episode Statistics; ONS, Office of National Statistics; THIN, The Health Improvement Network; O:E ratio, observed:expected ratio.

Notes:

a. Livingstone 2023 is a sub-cohort of Livingstone 2022.
b. Very serious risk of bias in the evidence contributing to the outcomes. More than 50% of the weight of the evidence came from studies at high risk of bias as per PROBAST.
c. Population is partially applicable due to study limited to adults from the general population that includes participants not at suspected risk or not at risk of fragility fracture (women < 65 and men < 75 years).

1

d. Very serious imprecision because 95% CI crosses 2 clinical decision thresholds (0.5 and 0.7) or serious imprecision because 95% CI crosses 1 clinical decision threshold (0.5 or 0.7).

e. Very serious inconsistency between internal and external validation results with non-overlapping 95% confidence intervals and point estimates either side of clinical decision threshold of 0.7.

f. Serious inconsistency between internal and external validation results with non-overlapping 95% confidence intervals.

Table 17: GRADE profile for discriminatory power of risk prediction tools to predict 10-year risk of hip fracture

Population Type of validation study Validation cohort, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	c-statistic (95% CI)	GRADE certainty
CFracture Men 30-99 <u>Internal prospective cohort</u> - CPRD GOLD, HES, ONS (Livingstone 2023[a]), N=894,910	1	Very serious[b]	Not serious	Serious[c]	Not serious	0.886 (0.877-0.895)	VERY LOW
CFracture Women 30-99 <u>Internal prospective cohort</u> - CPRD GOLD, HES, ONS (Livingstone 2023[a]), N=915,803	1	Very serious[b]	Not serious	Serious[c]	Not serious	0.914 (0.908-0.919)	VERY LOW
FRAX-UK without BMD	1	Very serious[b]	Not serious	Serious[c]	Not serious	0.83 (0.812-0.847)	VERY LOW

Population Type of validation study Validation cohort, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	c-statistic (95% CI)	GRADE certainty
Adults 40-90 External prospective cohort - CPRD with HES linkage (Klop 2016), N=24,227							
FRAX-UK without BMD Men 40-85 External prospective cohort - QResearch, v.20 (Hippisley-Cox 2009), N=424,336	1	Very serious[b]	Not serious	Serious[c]	Not serious	0.817 (0.807-0.827)	VERY LOW
FRAX-UK without BMD Women 40-85 External prospective cohort QResearch, v.20 (Hippisley-Cox 2009), N=454,499	1	Very serious[b]	Not serious	Serious[c]	Not serious	0.845 (0.840-0.850)	VERY LOW

Population Type of validation study Validation cohort, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	c-statistic (95% CI)	GRADE certainty
FRAX-UK without BMD Adults with COPD≥40 years-old External prospective cohort THIN (Akyea 2019), N=72,559	1	Very serious[b]	Not serious	Serious[c]	Not serious	0.761 (0.749-0.772)	VERY LOW
FRAX-UK without BMD Adults with RA, 40-90 External prospective cohort CPRD (Klop 2016), N=11,582	1	Very serious[b]	Not serious	Serious[c]	Not serious	0.78 (0.752-0.805)	VERY LOW
QFracture 2009 Men 30-85 Internal prospective cohort	2	Very serious[b]	Very serious[d]	Serious[c]	Not serious	Internal validation - 0.688 (0.684-0.692)	VERY LOW

Population Type of validation study Validation cohort, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	c-statistic (95% CI)	GRADE certainty
- QResearch v.20 (Hippisley-Cox 2009), N=633,764 External prospective cohort - THIN (Collins 2011), N=1,108,219						External validation - 0.739 (0.733-0.745)	
QFracture 2009 Women 30-85 Internal prospective cohort - QResearch v.20 (Hippisley-Cox 2009), N=642,153 External prospective cohort - THIN (Collins 2011), N=1,136,417	2	Very serious[b]	Serious[d]	Serious[c]	Not serious	Internal validation - 0.788 (0.786-0.790) External validation - 0.816 (0.813-0.819)	VERY LOW
QFracture 2012 Men 30-100 Internal prospective cohort	3	Very serious[b]	Not serious	Serious[c]	Not serious	Internal validation - 0.875 (0.868-0.883)	VERY LOW

Population Type of validation study Validation cohort, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	c-statistic (95% CI)	GRADE certainty
- QResearch v.32 (Hippisley-Cox 2012), N=778,810 <u>External prospective cohorts</u> - CPRD (Hippisley-Cox 2014), N=1,588,803 - CPRD GOLD, HES, ONS (Livingstone 2022), N=2,684,730						<u>External validation</u> - 0.872 (0.867-0.877) HC2014 - 0.888 (0.882-0.893) LG2022	
QFracture 2012 Women 30-100 <u>Internal prospective cohort</u> - QResearch v.32 (Hippisley-Cox 2012), N=804,563 <u>External prospective cohorts</u> - CPRD (Hippisley-Cox 2014), N=1,682,709 CPRD GOLD, HES, ONS (Livingstone 2022), N=2,747,409	3	Very serious[b]	Not serious	Serious[c]	Not serious	<u>Internal validation</u> - 0.893 (0.890-0.896) <u>External validation</u> - 0.890 (0.888-0.892) HC2014 - 0.918 (0.915-0.921) LG 2022	VERY LOW

Population Type of validation study Validation cohort, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	c-statistic (95% CI)	GRADE certainty
QFracture 2016 Adults with COPD≥40 years-old External prospective cohort - THIN (Akyea 2019), N=72,559	1	Very serious[b]	Not serious	Serious[c]	Not serious	- 0.761 (0.749-0.772)	VERY LOW

Abbreviations: COPD, chronic obstructive pulmonary disease; CPRD, Clinical Practice Research Datalink; FRAX, Fracture risk assessment tool; HES, Hospital Episode Statistics; ID, intellectual disabilities; ONS, Office of National Statistics; RA, rheumatoid arthritis; THIN, The Health Improvement Network; O:E ratio, observed:expected ratio.

Notes:

a. *Livingstone 2023 uses a sub-cohort of Livingstone 2022.*
b. *Very serious risk of bias in the evidence contributing to the outcomes. More than 50% of the weight of the evidence came from studies at high risk of bias as per PROBAST.*
c. *Population is partially applicable due to study limited to adults from the general population that includes participants not at suspected risk or not at risk of fragility fracture (women < 65 and men < 75 years).*
d. *Very serious inconsistency between internal and external validation results with non-overlapping 95% confidence intervals and point estimates either side of clinical decision threshold of 0.7.*

1.1.10. Economic evidence

Economic evidence related to risk assessment tools was sought as part of Evidence Review E. No included studies compared alternative risk prediction tools.

1.1.11. Economic model

This area was not prioritised for new cost-effectiveness analysis.

1.1.12. Unit costs

FRAX and QFracture have free online calculators. FRAXplus® adjustments require the user to purchase an annual service plan, for example €50 for an individual user (£43 using September 2025 HMRC exchange rates).

CFracture did not have an online calculator available at the time of guideline development.

1.2. The committee's discussion and interpretation of the evidence

1.2.1. The outcomes that matter most

The committee agreed that the clinical outcomes the risk prediction tools should predict were major osteoporotic (MOF) fracture and hip fracture. MOF was defined as hip, clinical vertebral, humerus and forearm in accordance with FRAX and QFracture (2012) risk prediction tools.

The following statistical outcomes were identified as relevant to assessing the performance of risk prediction tools:

- Overall fit: R^2 and Brier score
- Calibration: calibration plots and curves, calibration in the large, observed:expected (O:E) ratio, integrated calibration index
- Discrimination: AUC/c-statistic and D statistic for overall discrimination, sensitivity, specificity, and predictive values at specific thresholds
- Reclassification statistics (Net reclassification index)

The committee recognised that there are, and have been, many proposed measures of assessing the performance of risk prediction tools and that validation studies have often reported some (but not all) of the above measures. Although evidence was identified on all the listed UK-validated tools, reporting of the various performance measures was generally not comprehensive. Data on all available measures from the included studies were extracted but the committee decided to focus on the O:E ratio to assess calibration and AUC/c-statistic values to assess discriminatory power across all possible thresholds, as these were reported in most identified studies.

The committee focussed on assessing the calibration plots and associated estimated O:E ratio (a measure of how well on average the observed and predicted risks agree) as several of the included studies reported sufficient data to calculate this. AUC/c-statistic values and O:E ratios were either extracted from the studies or calculated from the available data or calibration plots in line with the methods detailed in Debray 2018. Other calibration measures such as calibration slope and calibration-in-the-large were largely not reported in most studies. Reclassification decisions are also important to compare the utility of the tools as they assess whether the new

prediction model improves on the old model. Similar to measures of calibration, most studies did not report reclassification measures.

There were few studies on each risk prediction tool and so meta-analysis of their overall calibration and discriminatory power was not possible.

Clinical decision thresholds

Clinical decision thresholds were set as default values for AUC, sensitivity, and specificity, indicating that a test would be recommended if 0.7 and above and not recommended if below 0.5 indicating that a test is no better than chance and therefore of no clinical use. The AUC describes the overall prognostic accuracy across the full range of possible thresholds. The following criteria were used for evaluating AUCs:

- ≤0.50: worse than chance
- 0.50–0.60: very poor
- 0.61–0.70: poor
- 0.71–0.80: moderate
- 0.81–0.90: good
- 0.91–1.00: excellent or perfect test.

1.2.2. The quality of the evidence

GRADE assessment was conducted on the discriminatory power of the fracture risk prediction tool using the reported c-statistics or AUC values as this was reported for most studies. All the risk prediction tools for the outcomes of MOF and HF were assessed as low to very low certainty. They were all downgraded for high risk of bias, using the PROBAST tool, mainly due to ‘measurement error’ in the various domains: information about predictor variables, fracture ascertainment, or statistical frameworks not accounting for competing risks (for example, QFracture).

Some discrimination outcomes were downgraded for directness and assessed as partially applicable because the majority of the studies were conducted in unselected populations with people below the age risk threshold (women below 65 and men below 75 years). This meant that studies included people not at risk or suspected risk of fragility fracture. Some outcomes were also downgraded for imprecision due to the 95% confidence intervals crossing 1 or 2 of the clinical decision thresholds (0.5 and 0.7).

Inconsistency between the internal validation study of the tool and external validation studies was assessed for the 2009 and 2012 versions of QFracture. There was serious inconsistency for prediction of MOF and HF by QFracture 2009 in women with non-overlapping 95%CIs. There was also very serious inconsistency in men for both MOF and HF outcomes as, in addition the AUC point estimates were on either side of the clinical decision thresholds; for QFracture 2012, there was serious inconsistency for prediction of MOF in both men and women.

One study (Livingstone 2023) used two-thirds of the same population cohort as another included study (Livingstone 2022) to develop CFracture and the remaining one-third of the cohort to internally validate CFracture and directly compare it to the performance of (that is, externally validate) QFracture 2012. Data on this direct comparison between CFracture and QFracture 2012 was not included in this review because data for the whole cohort (reported in Livingstone 2022) is already included and the results are likely to be similar.

1.2.3. Benefits and harms

Generally, the performance of the UK-validated tools in terms of both calibration and discrimination was acceptable for the estimation of an individual's 10-year risk of MOF and HF. The committee recognised that there were substantive differences in how the models underlying the UK-validated tools are constructed, with different statistical models and associated risk estimators, and different methods of fracture ascertainment, used. Risk prediction tools that account for competing risk of mortality, such as FRAX and CFracture, are in principle better models. This is because those that do not account for competing risk of mortality, such as QFracture, make the assumption that individuals lost to follow up (who are more likely to be older and have comorbidities) have the same fracture risk as those not lost to follow-up. This can lead to systematic overprediction of fracture risk depending on the frequency of competing risk events in a particular validation cohort. CFracture, a more recently constructed tool, is like QFracture as it was constructed using the same variables as QFracture and a similar development dataset. One feature of this tool is that, unlike QFracture, it accounts for competing mortality risk and appears to improve overall discriminatory performance of the tool with moderate or good AUC for MOF and good or excellent AUC for HF.

One study was identified for the IDFracture tool that is a risk prediction model for the identification of people with intellectual disabilities at risk of major osteoporotic fracture and hip fracture. The study was not included as it had not been peer-reviewed. The committee were aware that the study is awaiting publication and therefore have not made a research recommendation. The committee also noted that for the tool to be useful to NHS practitioners there would also need to be an online calculator for it.

FRAX-UK is widely available and accounts for competing risk of non-fracture mortality. However, the UK model has not been updated since 2008 (using fracture-incidence and mortality data from before 2000). Five studies of FRAX-UK in mostly small (<1000) sample high-fracture risk populations such as COPD were identified (Akyea 2019, Green 2024, Hippisley-Cox 2009, Ihama 2021, Klop 2016).

For the prediction of major osteoporotic fracture, three studies (Akyea 2016, Ihama 2021, Klop 2016) reported poor or moderate AUC values (ranging from 0.66 to 0.71). One study reported sensitivity, specificity, and predictive values of FRAX-UK with and without BMD using the NOGG age-specific thresholds, which showed very low sensitivity and PPV. This indicates that there is a substantive risk of false negatives (and therefore missed fractures). The committee acknowledged that although the discriminatory power of FRAX-UK to predict MOF outcomes was not excellent, it should be used alongside other factors to decide who should have a DXA scan.

For the prediction of hip fracture, three studies (Akyea 2019, Hippisley-Cox 2009, Klop 2016) showed that FRAX-UK had overall good or moderate discriminatory power with AUC values ranging from 0.76 to 0.85 in the general primary care population. The relatively narrow 95%CIs reflect the large size of the studies. One study in adults with COPD from the primary care population (Akyea 2019) estimated that although its sensitivity using a 3% fracture risk threshold was above the threshold for recommendation, it had low specificity and would therefore lead to a substantial number of false positives and patients receiving unnecessary further assessment.

No studies were identified on FRAXplus® adjustments which unlike the online FRAX calculator is not freely available online.

QFracture 2012 was shown to be well validated in 2 independent general primary care cohorts (Hippisley-Cox 2014, Livingstone 2022) and in general, it appears to perform better than the 2009 version. Performance was excellent with high values in all performance categories. The AUC values ranged from moderate to good for major osteoporotic fracture risk (0.74 – 0.82) and good to excellent (0.87-0.92) for hip fracture outcomes. However, the committee recognised that the model does not account for the competing risks of non-fracture mortality.

Calibration performance in these cohorts suggest that when more robust fracture ascertainment methods are used (such as linked hospital or mortality data), the tool tended to underestimate MOF and HF fracture risk in the cohort. For example, although calibration was improved for predicting 10-year risk of MOF in men when competing risks of non-fracture mortality were accounted for (for example, reducing O:E ratio from 1.82 to 1.48 in men), QFracture still substantially underestimated fracture risk. Calibration performance for QFracture 2012 in women was slightly better for women than for men in estimating MOF and HF risk, although it still underestimated fracture risk (Livingstone 2022).

Generally, the committee agreed that the lack of comprehensive reporting of the performance measures (that is, overall fit, calibration, discrimination, and reclassification) makes it difficult to assess the included UK-validated risk prediction tools.

QFracture 2016 is currently in use but there are few validation studies: the internal validation study is available but has not been published, and the tool still does not account for competing risks of non-fracture cause mortality. The two studies (Akyea 2019, Ihama 2021) conducted in high-risk populations reported AUC values of 0.74 and 0.61 for major osteoporotic fracture and 0.76 for hip fracture. The discriminatory power ranged from poor to moderate. Overall fit and calibration measures were not reported so the committee were not able to fully assess its benefits and harms.

CFracture accounts for competing risks of non-fracture mortality and had similar overall discriminatory power to QFracture 2012 for both MOF and HF in one internal validation study (approximately 1.7m people) (Livingstone 2023). Overall, CFracture performed better at estimating HF rather than MOF risk with potentially excellent calibration and discriminatory power in women (AUC values ranged from 0.74 and 0.81 for MOF and 0.89 and 0.91 for HF in men and women respectively).

1.2.4. Conclusions and committee experiences

The committee's overall approach to the risk assessment pathway in people identified as at risk or suspected risk of fracture is discussed in Evidence review E. Currently, using FRAX-UK or QFracture is recommended when using a risk prediction tool to estimate an individual's risk of MOF and HF for most age groups in the previous NICE guideline on osteoporosis (see Supporting Document G: NICE CG146 Osteoporosis Full Guideline and Appendices). The committee agreed that none of the evidence identified in this review and alongside evidence reviews D and E merited any change to this recommendation.

The evidence for both FRAX-UK and QFracture tools were similarly calibrated and had relatively similar discriminatory power and therefore, did not clearly favour one tool over the other. Since the estimated risk calculated by these tools for an individual are not commensurate, the committee emphasised that the same risk prediction tool should be used to allow consistent monitoring across time from baseline. The committee also recognised that since QFracture does not include BMD as a risk factor, clinicians wishing to adjust estimated risk using this information will need to

use the UK version of FRAX, which can incorporate measurements taken at the femoral neck. The online Qfracture calculator uses the latest version of the tool (2016) and although there is limited evidence for this version it was agreed that it is likely to have similar discrimination to the 2012 version.

The committee also noted that FRAX-UK is shorter than QFracture with the former requiring only 11 fields to be completed (12 if BMD is included) compared to 24 (26 if BMI is included) for the latter and that, in some regions of the country, FRAX is used to determine eligibility for DXA scan. A recommendation was made to highlight that FRAX and QFracture assess risk differently because they do not include all the same risk factors in their models and were developed in different cohorts with different age groups (see **Table 2**). The committee emphasised that the risk prediction tools can return different estimates for an individual and that care needs to be taken when assessing their clinical risk profile.

Important risk factors that are treated differently in the tools include high alcohol intake, family history of osteoporosis, secondary causes of osteoporosis, current medication use, and living in a care home. For example, alcohol intake is defined using 5 categories of alcohol intake in QFracture whilst FRAX has a yes or no question about consuming 3 or more units of alcohol per day. The committee recommended to do a full clinical risk assessment alongside the risk prediction tool because QFracture and FRAX do not include all factors associated with an increased risk of fracture (for example, if a person is taking medicines associated with accelerated bone loss such as aromatase inhibitors or androgen deprivation therapy).

1.2.4.1. Research recommendation

Although the CFracture model and associated development and validation data is publicly available, there is no available calculator, and the committee therefore could not recommend it. Given the promising results from the available internal validation study, the committee agreed that a research recommendation should be made for external validation studies comparing the performance of CFracture and the latest available versions of QFracture and FRAX-UK.

1.2.5. Cost effectiveness and resource use

None of the identified cost-effectiveness analyses in Review E compared risk assessment with QFracture and FRAX.

The committee agreed that in terms of the fragility fracture risk assessment itself, the cost of using FRAX (without BMD) and QFracture would be similar as they would take similar time to complete, and both have freely available online calculators. It was agreed they could both be completed as part of a 15-minute GP appointment to discuss risk factors and patient history. It was noted that risk assessment may also take place in secondary care, for example if a patient is referred to a fracture liaison service or during a hospital admission.

Downstream resource use would depend on the rules applied in conjunction with the risk assessment in terms of who goes on to have a BMD assessment and treatment. In current practice, different rules are often applied with QFracture and FRAX as QFracture is often used in conjunction with SIGN guidance for DXA and treatment initiation and FRAX is generally used in conjunction with NOGG guidance which have different criteria for BMD assessment and treatment. However, the committee agreed that the same principles should be applied regarding BMD assessment and treatment

initiation irrespective of risk prediction tool used and that downstream resource use was also considered likely to be similar in this case.

The committee noted that risk can be estimated with or without BMD with FRAX. The committee recommendations for initial assessment to determine eligibility of DXA necessarily relate to risk calculation without BMD. The committee did not specifically recommend that risk is recalculated following BMD assessment although noted that clinicians may choose to do this in some cases. The committee agreed that there would be a follow-up appointment after DXA and that if needed risk could be recalculated during this or it may be done as part of the DXA scan and included in the report. It was therefore considered unlikely to result in a significant resource use difference between the tools even when done.

The committee noted that while the online FRAX calculator is free to use, use of FRAXplus® adjustments incurs additional costs. However, the committee did not make recommendations related to use of these adjustments due to a lack of clinical evidence.

The committee discussed whether the fact that different risk prediction tools can return different estimates for an individual (discussed in previous section above) could mean that the choice of tool might result in different numbers of people eligible for DXA and so lead to differences in resource use. The committee agreed that while sometimes different people may be considered high risk with different tools the overall numbers considered eligible for DXA was likely to be similar, if the same threshold was used, and this was unlikely to lead to substantial differences in downstream resource use between the use of either of the two tools.

The recommendation for use of either QFracture or FRAX when using a fracture risk prediction tool is not a change from the previous NICE guideline. The committee advised that FRAX is currently more widely used in England currently as most people use the NOGG guideline which uses FRAX, and FRAX score is sometimes built into DXA referral processes. The committee agreed however that QFracture is also used. QFracture is used most commonly in Scotland due to the SIGN guideline that states a preference for this tool. The recommendation about choice of risk prediction tool is not expected to result in a significant resource impact to the NHS.

1.2.6. Other factors the committee took into account

The committee noted that people with learning difficulties may find completing the forms difficult depending on the severity of their condition. However, the forms are generally completed by healthcare professionals rather than the individual themselves, but extended GP appointment time may be needed. People with learning disabilities or cognitive impairment may be less able to provide accurate answers in relation to risk factors that are not recorded in the medical records (for example, parental hip fracture).

1.2.7. Recommendations supported by this evidence review

This evidence review supports recommendations 1.3.1-1.3.9 and the research recommendation on the validity of CFracture risk prediction tool for predicting the risk of fragility fractures in adults, including those who have had a previous fragility fracture. There is overlap between evidence reviews and recommendations from evidence reports C, D and E.

1 # 1.3. References

2 ## 1.3.1. Clinical

Akyea, Ralph Kwame; McKeever, Tricia M; Gibson, Jack et al. (2019) Predicting fracture risk in patients with chronic obstructive pulmonary disease: a UK-based population-based cohort study. BMJ open 9(4): e024951

Collins, Gary S; Mallett, Susan; Altman, Douglas G (2011) Predicting risk of osteoporotic and hip fracture in the United Kingdom: prospective independent and external validation of QFractureScores. BMJ (Clinical research ed.) 342: d3651

Green, Olivia; Raju, Suneil A; Shiha, Mohamed G et al. (2024) Clinical utility of the fracture risk assessment tool (FRAX) in biopsy-confirmed coeliac disease. Scandinavian journal of gastroenterology 59(9): 1049-1054

Hippisley-Cox, Julia and Coupland, Carol (2012) Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study. BMJ (Clinical research ed.) 344: e3427

Hippisley-Cox, Julia and Coupland, Carol (2009) Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores. BMJ (Clinical research ed.) 339: b4229

Hippisley-Cox, Julia; Coupland, Carol; Brindle, Peter (2014) The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study. BMJ open 4(8): e005809

Ihama, F; Pandyan, A; Roffe, C (2021) Assessment of fracture risk tools in care home residents: a multi-centre observational pilot study. European geriatric medicine 12(1): 79-89

Klop, Corinne; de Vries, Frank; Bijlsma, Johannes W J et al. (2016) Predicting the 10-year risk of hip and major osteoporotic fracture in rheumatoid arthritis and in the general population: an independent validation and update of UK FRAX without bone mineral density. Annals of the rheumatic diseases 75(12): 2095-2100

Livingstone, Shona J; Guthrie, Bruce; McMinn, Megan et al. (2023) Derivation and validation of the CFracture competing risk fracture prediction tool compared with QFracture in older people and people with comorbidity: a population cohort study. The lancet. Healthy longevity 4(1): e43-e53

Livingstone, Shona J; Morales, Daniel R; McMinn, Megan et al. (2022) Effect of competing mortality risks on predictive performance of the QFracture risk prediction tool for major osteoporotic fracture and hip fracture: external validation cohort study in a UK primary care population. BMJ medicine 1(1): e000316

1.3.2. Economic

Economic evidence related to risk assessment tools was sought as part of Evidence review E.

1.3.3. Other

1. Debray, Thomas PA; Damen, Johanna AAG; Riley, RD et al. (2018) A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Statistical Methods in Medical Research 28(9):2768-2786.

Appendices

Appendix A Review protocols

A.1.1 Review protocol: What is the validity of risk prediction tools for predicting the risk of fragility fractures in adults, including those who have had a previous fragility fracture?

Field	Content
Review title	Fragility fracture risk prediction tools and bone assessment methods to predict fragility fracture
Review question	3.1a What is the validity of risk prediction tools for predicting the risk of fragility fractures in adults, including those who have had a previous fragility fracture?
Objective	Fractures associated with osteoporosis, often described as 'fragility fractures,' typically result from a low impact injury such as a fall from standing height or less which would otherwise not be expected to result in a fracture. Fragility fractures can occur spontaneously with no history of injury and most vertebral fractures do not result from a fall but are precipitated by an activity involving lifting, twisting, or bending. This review will update NICE guideline CG146 and will evaluate (i) the validity of risk prediction tools in the same or different population/setting used to develop model, and (ii) the accuracy of bone assessment methods in adults (and associated optimum thresholds), for predicting the risk of fragility fracture in adults, including those who have had a previous fragility fracture.
Searches	The following databases (from inception) will be searched: - Embase - MEDLINE - Epistemonikos Searches will be restricted by: - English language studies - Human studies Other searches: - Reference searching

	- Citation searching - Inclusion lists of systematic reviews The searches may be re-run 6 weeks before the final committee meeting and further studies retrieved for inclusion if relevant. The full search strategies will be published in the final review. Medline search strategy to be quality assured using the PRESS evidence-based checklist (see methods chapter for full details).
Condition or domain being studied	Fragility fracture
Population	**Inclusion:** Adults (18 years and older) who are at suspected risk of fragility fracture (people with or at risk of primary or secondary osteoporosis or have had a previous fragility fracture) **Exclusion**: Children and young people less than 18 years

Risk prediction tool/bone assessment method	The following multivariable risk prediction tools for major osteoporotic fracture (MOF) or hip fracture (HF), which have been validated in a UK-population, will be included:
	• CFracture
	• FRAX®-UK/FRAXplus®-UK
	  o Without bone mineral density assessment (BMD)
	  o With BMD
	  o With BMD and trabecular bone score (TBS)
	• FRAX with NOGG thresholds
	• IDFracture
	• QFracture
	Strata: Version or iteration of risk prediction tool; Type of fracture (MOF, HF).
	Note: This is an amendment to the initial protocol, undertaken after the initiation of data analysis, to clarify the following risk tools and bone assessment methods:
	- addition of IDFracture tool

	- clarification that FRAX with additional analyses refers to FRAX with NOGG thresholds.
Target condition	Fragility fracture. • Major osteoporotic fracture (MOF) • Hip fracture (HF) For risk prediction tools, timing is for 5- or 10-year risk of MOF or HF
Types of study to be included	**Risk prediction tools** **Inclusion:** • External validation cohort studies • Internal-external cross validation cohort studies • Internal validation cohort studies of the included fragility risk prediction tools The original internal validation studies of each risk prediction model will be included to enable comparison to the relevant identified external validation studies. External validation studies may be in the same populations and setting used in the development of the prediction model, or in different populations or settings. Such studies may also compare more than one fragility fracture risk prediction tool. **Exclusion** • For FRAX®-UK/FRAXplus®, validation studies not conducted in the UK • Studies using machine learning algorithms, polygenic risk scores, or radiomic models will be excluded Note: This is an amendment to the initial protocol, undertaken after the initiation of data analysis, to clarify that only UK validation studies of risk tools will be included.
Other exclusion criteria	• Non-English language studies • Conference abstracts
Context	All settings
Primary outcomes (critical outcomes)	The validity of risk prediction tools for fragility fracture will be evaluated using the following measures: • Overall fit

	○ R^2 statistic (for continuous outcomes) ○ Brier score (for binary/time-to-event outcomes) • Calibration (agreement between observed and predicted values) ○ Calibration-in-the-large ○ Observed/expected ratio ○ Calibration plots and curves ○ Integrated calibration index • Overall discrimination for binary or to-event outcomes (fracture v no fracture) ○ c-statistic/Area under the curve [AUC]) for binary outcomes ○ Harrell's Concordance (C) index or Royston's D statistic for time-to-event outcomes • Reclassification (e.g. Net Reclassification Index) • Discrimination at specific threshold ○ Predictive values ○ Sensitivity and specificity All outcomes are considered equally important for decision making and therefore have all been rated as critical.
Data extraction (selection and coding)	All references identified by the searches and from other sources will be uploaded into EPPI R5 and de-duplicated. Titles and abstracts of the retrieved citations will be screened to identify studies that potentially meet the inclusion criteria outlined in the review protocol. Dual sifting will be performed on at least 10% of records; 90% agreement is required. Disagreements will be resolved via discussion between the two reviewers, and consultation with senior staff if necessary.

	Full versions of the selected studies will be obtained for assessment. Studies that fail to meet the inclusion criteria once the full version has been checked will be excluded at this stage. Each study excluded after checking the full version will be listed, along with the reason for its exclusion. A standardised form will be used to extract data from studies. The following data will be extracted: study details (reference, country where study was carried out, type and dates), participant characteristics, inclusion and exclusion criteria, details of the interventions if relevant, setting and follow-up, relevant outcome data, and source of funding. One reviewer will extract relevant data into a standardised form, and this will be quality assessed by a senior reviewer.	
Risk of bias (quality) assessment	Risk of bias will be assessed using the appropriate checklist as described in Developing NICE guidelines: the manual: • PROBAST for risk prediction tool studies A second reviewer will quality assure 10% of the critical appraisal assessments. Discrepancies will be identified and resolved through discussion (with a third party where necessary).	
Strategy for data synthesis	Heterogeneity between the studies in effect measures will be assessed using the I^2 statistic and visually inspected. An I^2 value greater than 50% will be considered indicative of substantial heterogeneity. Sensitivity analyses will be conducted based on pre-specified subgroups using stratified meta-analysis to explore the heterogeneity in effect estimates. If this does not explain the heterogeneity, the results will be presented pooled using random effects. - GRADEpro will be used to assess the quality of evidence for each outcome, taking into account individual study quality and the meta-analysis results. The 4 main quality elements (risk of bias, indirectness, inconsistency, and imprecision) will be appraised for each outcome. Publication bias will be considered with the guideline committee, and if suspected will be tested for when there are more than 5 studies for that outcome. The risk of bias across all available evidence was evaluated for each outcome using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group http://www.gradeworkinggroup.org/ - Where meta-analysis is not possible, data will be presented and quality assessed individually per outcome. - WinBUGS will be used for network meta-analysis, if possible, given the data identified.	
Analysis of sub-groups	Subgroups that will be investigated if heterogeneity is present: • People who have received treatment that affects bone density; People who have not received treatment that affects bone density)	
	☒	Intervention

Type and method of review	☐	Diagnostic	
	☒	Prognostic	
	☐	Qualitative	
	☐	Epidemiologic	
	☐	Service Delivery	
	☒	Other – Risk tool review	
Language	English		
Country	England		
Anticipated or actual start date	2023		
Anticipated completion date	November 2025		
Stage of review at time of this submission	Review stage	Started	Completed
	Preliminary searches	☑	☑
	Piloting of the study selection process	☑	☑
	Formal screening of search results against eligibility criteria	☑	☑
	Data extraction	☑	☑
	Risk of bias (quality) assessment	☑	☑
	Data analysis	☑	☑
Named contact	5a. Named contact		

	Guideline Development Team NGC 5b Named contact e-mail Carlos.Sharpin@nice.org.uk 5e Organisational affiliation of the review National Institute for Health and Care Excellence (NICE)
Review team members	From NICE: Carlos Sharpin [Guideline lead] Julie Neilson [Senior research fellow] Clare Jones [Senior technical analyst] Annette Chalker [Technical analyst] Linyun Fou [Technical analyst] Kate Lovibond [Senior Health economist] Muksitur Rahman [Health economist] Sarah Glover [Information specialist] Stephen Deed [Information specialist] Claire Sloan [Information specialist]
Funding sources/sponsor	Development of this systematic review is being funded by NICE.
Conflicts of interest	All guideline committee members and anyone who has direct input into NICE guidelines (including the evidence review team and expert witnesses) must declare any potential conflicts of interest in line with NICE's code of practice for declaring and dealing with conflicts of interest. Any relevant interests, or changes to interests, will also be declared publicly at the start of each guideline committee meeting. Before each meeting, any potential conflicts of interest will be considered by the guideline committee Chair and a senior member of the development team. Any decisions to exclude a person from all or part of a meeting will be documented. Any changes to a member's declaration of interests will be recorded in the minutes of the meeting. Declarations of interests will be published with the final guideline.

Collaborators	Development of this systematic review will be overseen by an advisory committee who will use the review to inform the development of evidence-based recommendations in line with section 3 of Developing NICE guidelines: the manual. Members of the guideline committee are available on the NICE website: http://www.nice.org.uk/guidance/indevelopment/gid-ng10216	
Other registration details	N/A	
Reference/URL for published protocol	N/A	
Dissemination plans	NICE may use a range of different methods to raise awareness of the guideline. These include standard approaches such as: - notifying registered stakeholders of publication - publicising the guideline through NICE's newsletter and alerts - issuing a press release or briefing as appropriate, posting news articles on the NICE website, using social media channels, and publicising the guideline within NICE.	
Keywords	Assessment tool: accuracy; adults; bone assessment; bone mineral density (BMD); calibration; CFracture; computed tomography; dual-X-ray absorptiometry (DEXA, DXA); fragility fracture; fracture risk; FRAX; hip fracture; IDFracture; imaging; prediction tool; osteoporosis; hip fracture; osteoporotic fracture; QFracture; quantitative computed tomography (QCT); quantitative ultrasound (QUS); risk prediction; trabecular bone score; validation; X-ray.	
Details of existing review of same topic by same authors	Overview | Osteoporosis: assessing the risk of fragility fracture | Guidance | NICE	
Current review status	☐	Ongoing
	☒	Completed but not published
	☐	Completed and published
	☐	Completed, published, and being updated
	☐	Discontinued
Additional information	N/A	

Details of final publication	www.nice.org.uk

1

Appendix B Literature search strategies

The literature searches for this review are detailed below and complied with the methodology outlined in Developing NICE guidelines: the manual.(NICE2014) For more information, please see the Methodology review published as part of the accompanying documents for this guideline.

B.1 Clinical search literature search strategy

Searches were constructed using a PICO framework where population (P) terms were combined with Intervention (I) and in some cases Comparison (C) terms. Outcomes (O) are rarely used in search strategies as these concepts may not be indexed or described in the title or abstract and are therefore difficult to retrieve. Search filters were applied to the search where appropriate.

Q3.1a What is the validity of risk prediction tools for predicting the risk of fragility fractures in adults, including those who have had a previous fragility fracture?

Table 18: Database parameters, filters and limits applied

Database	Dates searched	Search filter used
Medline (OVID)	1946 – 15 November 2024	Prognostic studies Systematic reviews Exclusions (animal studies, letters, comments, editorials, case studies/reports) English language
Embase (OVID)	1974 – 15 November 2024	Prognostic studies Systematic reviews Exclusions (animal studies, letters, comments, editorials, case studies/reports) English language
The Cochrane Library (Wiley)	Cochrane Reviews to 202 Issue 11 of 12 CENTRAL to 2024 Issue 11 of 12	Exclusions (clinical trials, conference abstracts)
Epistemonikos (The Epistemonikos Foundation)	Inception to 15 November 2024	Systematic review studies Exclusions (Cochrane reviews) English language

Medline (Ovid) search terms

1	exp Osteoporosis/
2	(osteopor* or osteo-por* or osteop?eni* or osteo-p?eni*).tw,kf.
3	((age-relat* or agerelat* or perimenopaus* or peri-menopaus* or postmenopaus* or post-menopaus* or menopaus* or pathologic*) adj4 bone* adj4 (los* or mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or demineral* or strength* or quality or quantit*)).tw.
4	((abnormal* or secondary or early or prematur*) adj4 bone* adj4 (los* or reduc* or mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or strength* or quality or quantit*)).tw.
5	((low* or reduc* or decreas* or los*) adj4 bone* adj4 (mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or strength* or quality or quantit*)).tw.
6	((age-relat* or agerelat* or perimenopaus* or peri-menopaus* or postmenopaus* or post-menopaus* or menopaus* or pathologic*) adj4 BMD).tw.
7	((low* or los* or reduc* or decreas* or abnormal* or secondary) adj4 BMD).tw.
8	(bone* adj4 (deteriorat* or weak* or fragil* or decalc* or brittle* or atroph*)).tw.
9	((trabecula* or cancellous) adj4 (loss* or thin* or reduc* or decreas* or deteriorat* or low* or abnormal*)).tw.
10	((age-relat* or agerelat* or perimenopaus* or peri-menopaus* or postmenopaus* or post-menopaus* or menopaus* or pathologic*) adj4 skeletal adj4 (los* or mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or demineral* or strength* or quality or quantit* or decalc* or atroph*)).tw.
11	((abnormal* or secondary or early or prematur*) adj4 skeletal* adj4 (los* or reduc* or mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or strength* or quality or quantit* or atroph*)).tw.
12	((low* or reduc* or decreas* or los*) adj4 skeletal adj4 (mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or strength* or quality or quantit*)).tw.
13	Bone Diseases, Metabolic/
14	Osteoporotic Fractures/
15	(fragil* adj4 (fracture or fractures)).tw.
16	((low-impact* or low-energy or low-trauma* or insufficien*) adj4 fracture*).tw.
17	((risk* or frequen* or inciden* or suscept* or suspect* or predict* or prevent* or stop*) adj4 fracture*).tw.
18	((recurrent or recurring or repeat* or history or chronic or previous or prior or habitual) adj4 fracture*).tw.
19	refracture*.tw.

20	or/1-19
21	exp Densitometry/
22	(densitometr* or BMD-test* or BMD-tool* or densimetr*).tw.
23	(bone adj4 mineral adj4 dens* adj4 test*).tw.
24	(bone adj4 mineral adj4 dens* adj4 tool*).tw.
25	Absorptiometry, Photon/
26	(absorptiometr* adj4 (dpx* or dual-energ* or dual-photon* or photon*)).tw.
27	X-Rays/
28	(x-ray* or xray*).tw.
29	((grenz* or roentgen*) adj4 ray*).tw.
30	(x-radiation* or xradiation*).tw.
31	(DXA* or DEXA).tw.
32	(FRAX or FRAXTM or Qfracture* or Q-fracture* or Cfracture* or C-Fracture*).tw.
33	(fracture* adj2 risk adj2 assess* adj2 tool*).tw.
34	(QCT* or pQCT* or HR-pQCT* or HRpQCT* or PCD-CT* or PCDCT* or SR-MUCT* or SRMUCT* or HRclinCT* HRclin-CT* or HR-clin-CT* or HR-clinCT*).tw.
35	(QUS or PEUS or P-EU or P-EUS or PEQUS).tw.
36	or/21-35
37	Tomography, X-Ray Computed/
38	(cat scan* or ct scan* or cine ct or cine-ct or tomodensitomet*).tw.
39	((computed or computer assisted or computeriz* or computeris* or electron beam* or axial*) adj4 tomograph*).tw.
40	Four-Dimensional Computed Tomography/
41	(4d ct or 4dct or 4-dimensional CT or four dimensional CT).tw.
42	exp Tomography, Spiral Computed/
43	((helical or spiral) adj4 ct*).tw.
44	exp Ultrasonography/
45	(ultrasound* or ultra-sound* or ultrason* or sonograph* or echograph* or echotomograph*).tw.
46	(bindex* or echolight*).tw.
47	or/37-46

48	(quantitative* or asynchronous or high-res* or highres or photon-count* or photoncount* or pulse-echo* or pulseecho* or pulsecho*).tw.
49	47 and 48
50	36 or 49
51	20 and 50
52	predict.ti.
53	(validat* or rule*).ti,ab.
54	(predict* and (outcome* or risk* or model*)).ti,ab.
55	((history or variable* or criteria or scor* or characteristic* or finding* or factor*) and (predict* or model* or decision* or identif* or prognos*)).ti,ab.
56	decision*.ti,ab. and Logistic models/
57	(decision* and (model* or clinical*)).ti,ab.
58	(prognostic and (history or variable* or criteria or scor* or characteristic* or finding* or factor* or model*)).ti,ab.
59	(stratification or discrimination or discriminate or c statistic or "area under the curve" or AUC or calibration or indices or algorithm or multivariable).ti,ab.
60	ROC curve/
61	or/52-60
62	Meta-Analysis/
63	exp Meta-Analysis as Topic/
64	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
65	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
66	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
67	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
68	(search* adj4 literature).ab.
69	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
70	cochrane.jw.
71	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
72	or/62-71
73	61 or 72

74	51 and 73
75	animals/ not humans/
76	74 not 75
77	limit 76 to (letter or historical article or comment or editorial or news or case reports)
78	76 not 77
79	limit 78 to english language

1

2 **Embase (Ovid) search terms**

1	exp Osteoporosis/
2	exp Osteopenia/
3	(osteopor* or osteo-por* or osteop?eni* or osteo-p?eni*).tw,kf.
4	((age-relat* or agerelat* or perimenopaus* or peri-menopaus* or postmenopaus* or post-menopaus* or menopaus* or pathologic*) adj4 bone* adj4 (los* or mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or demineral* or strength* or quality or quantit*)).tw.
5	((abnormal* or secondary or early or prematur*) adj4 bone* adj4 (los* or reduc* or mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or strength* or quality or quantit*)).tw.
6	((low* or reduc* or decreas* or los*) adj4 bone* adj4 (mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or strength* or quality or quantit*)).tw.
7	((age-relat* or agerelat* or perimenopaus* or peri-menopaus* or postmenopaus* or post-menopaus* or menopaus* or pathologic*) adj4 BMD).tw.
8	((low* or los* or reduc* or decreas* or abnormal* or secondary) adj4 BMD).tw.
9	(bone* adj4 (deteriorat* or weak* or fragil* or decalc* or brittle* or atroph*)).tw.
10	((trabecula* or cancellous) adj4 (loss* or thin* or reduc* or decreas* or deteriorat* or low* or abnormal*)).tw.
11	((age-relat* or agerelat* or perimenopaus* or peri-menopaus* or postmenopaus* or post-menopaus* or menopaus* or pathologic*) adj4 skeletal adj4 (los* or mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or demineral* or strength* or quality or quantit* or decalc* or atroph*)).tw.
12	((abnormal* or secondary or early or prematur*) adj4 skeletal* adj4 (los* or reduc* or mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or strength* or quality or quantit* or atroph*)).tw.
13	((low* or reduc* or decreas* or los*) adj4 skeletal adj4 (mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or strength* or quality or quantit*)).tw.

14	metabolic bone disease/ or exp bone demineralization/
15	fragility fracture/
16	(fragil* adj4 (fracture or fractures)).tw.
17	((low-impact* or low-energy or low-trauma* or insufficien*) adj4 fracture*).tw.
18	((risk* or frequen* or inciden* or suscept* or suspect* or predict* or prevent* or stop*) adj4 fracture*).tw.
19	((recurrent or recurring or repeat* or history or chronic or previous or prior or habitual) adj4 fracture*).tw.
20	refracture*.tw.
21	or/1-20
22	Bone densitometry/
23	(densitometr* or BMD-test* or densimetr*).tw.
24	(bone adj4 mineral adj4 dens* adj4 test*).tw.
25	(bone adj2 mineral adj2 dens* adj2 tool*).tw.
26	Photon absorptiometry/
27	(absorptiometr* adj4 (dpx* or dual-energ* or dual-photon* or photon*)).tw.
28	X ray/ or dual energy X ray absorptiometry/
29	(x-ray* or xray*).tw.
30	((grenz* or roentgen*) adj4 ray*).tw.
31	(x-radiation* or xradiation*).tw.
32	(DXA* or DEXA).tw.
33	FRAX tool/ or Qfracture/
34	(FRAX or FRAXTM or Qfracture* or Q-fracture* or Cfracture* or C-Fracture*).tw.
35	(fracture* adj2 risk adj2 assess* adj2 tool*).tw.
36	(QCT* or pQCT* or HR-pQCT* or HRpQCT* or PCD-CT* or PCDCT* or SR-MUCT* or SRMUCT* or HRclinCT* HRclin-CT* or HR-clin-CT* or HR-clinCT*).tw.
37	(QUS or PEUS or P-EU or P-EUS or PEQUS).tw.
38	or/22-37
39	X-ray computed tomography/
40	(cat scan* or ct scan* or cine ct or cine-ct or tomodensitomet*).tw.
41	((computed or computer assisted or computeriz* or computeris* or electron beam* or axial*) adj4 tomograph*).tw.

42	Four dimensional computed tomography/
43	(4d ct or 4dct or 4-dimensional CT or four dimensional CT).tw.
44	exp Tomography, Spiral Computed/
45	((helical or spiral) adj4 ct*).tw.
46	exp echography/
47	(ultrasound* or ultra-sound* or ultrason* or sonograph* or echograph* or echotomograph*).tw.
48	(bindex* or echolight*).tw.
49	or/394-48
50	(quantitative* or asynchronous or high-res* or highres or photon-count* or photoncount* or pulse-echo* or pulseecho* or pulsecho*).tw.
51	49 and 50
52	38 or 51
53	21 and 52
54	predict.ti.
55	(validat* or rule*).ti,ab.
56	(predict* and (outcome* or risk* or model*)).ti,ab.
57	((history or variable* or criteria or scor* or characteristic* or finding* or factor*) and (predict* or model* or decision* or identif* or prognos*)).ti,ab.
58	decision*.ti,ab. and Statistical model/
59	(decision* and (model* or clinical*)).ti,ab.
60	(prognostic and (history or variable* or criteria or scor* or characteristic* or finding* or factor* or model*)).ti,ab.
61	(stratification or discrimination or discriminate or c statistic or "area under the curve" or AUC or calibration or indices or algorithm or multivariable).ti,ab.
62	Receiver operating characteristic/
63	or/54-62
64	systematic review/
65	meta-analysis/
66	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
67	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
68	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.

69	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
70	(search* adj4 literature).ab.
71	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
72	cochrane.jw.
73	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
74	or/64-73
75	63 or 74
76	53 and 75
77	(conference abstract* or conference review or conference paper or conference proceeding).db,pt,su.
78	76 not 77
79	nonhuman/ not human/
80	78 not 79
81	(letter or editorial).pt.
82	80 not 81
83	limit 82 to english language

Cochrane Library (Wiley) search terms

#1	MeSH descriptor: [Osteoporosis] explode all trees
#2	((osteopor* or osteo-por* or osteopeni* or osteo-peni* or osteopaeni* or osteo-paeni*)):ti,ab,kw
#3	(((age-relat* or agerelat* or perimenopaus* or peri-menopaus* or postmenopaus* or post-menopaus* or menopaus* or pathologic*) near/4 bone* near/4 (los* or mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or demineral* or strength* or quality or quantit*))):ti,ab,kw
#4	(((abnormal* or secondary or early or prematur*) near/4 bone* near/4 (los* or reduc* or mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or strength* or quality or quantit*))):ti,ab,kw
#5	(((low* or reduc* or decreas* or los*) near/4 bone* near/4 (mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or strength* or quality or quantit*))):ti,ab,kw
#6	(((age-relat* or agerelat* or perimenopaus* or peri-menopaus* or postmenopaus* or post-menopaus* or menopaus* or pathologic*) near/4 BMD)):ti,ab,kw
#7	(((low* or los* or reduc* or decreas* or abnormal* or secondary) near/4 BMD)):ti,ab,kw

#8	((bone* near/4 (deteriorat* or weak* or fragil* or decalc* or brittle* or atroph*))):ti,ab,kw
#9	(((trabecula* or cancellous) near/4 (loss* or thin* or reduc* or decreas* or deteriorat* or low* or abnormal*))):ti,ab,kw
#10	(((age-relat* or agerelat* or perimenopaus* or peri-menopaus* or postmenopaus* or post-menopaus* or menopaus* or pathologic*) near/4 skeletal near/4 (los* or mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or demineral* or strength* or quality or quantit* or decalc* or atroph*))):ti,ab,kw
#11	(((abnormal* or secondary or early or prematur*) near/4 skeletal* near/4 (los* or reduc* or mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or strength* or quality or quantit* or atroph*))):ti,ab,kw
#12	(((low* or reduc* or decreas* or los*) near/4 skeletal near/4 (mass or architectur* or microarchitectur* or micro-architectur* or dens* or mineral* or content or strength* or quality or quantit*))):ti,ab,kw
#13	MeSH descriptor: [Bone Diseases, Metabolic] this term only
#14	MeSH descriptor: [Osteoporotic Fractures] this term only
#15	((fragil* near/4 (fracture or fractures))):ti,ab,kw
#16	(((low-impact* or low-energy or low-trauma* or insufficien*) near/4 fracture*)):ti,ab,kw
#17	(((risk* or frequen* or inciden* or suscept* or suspect* or predict* or prevent* or stop*) near/4 fracture*)):ti,ab,kw
#18	(((recurrent or recurring or repeat* or history or chronic or previous or prior or habitual) near/4 fracture*)):ti,ab,kw
#19	(refracture*):ti,ab,kw
#20	{or #1-#19}
#21	MeSH descriptor: [Densitometry] explode all trees
#22	((densitometr* or BMD-test* or densimetr*)):ti,ab,kw
#23	((bone near/4 mineral near/4 dens* near/4 test*)):ti,ab,kw
#24	((bone NEAR/4 mineral NEAR/4 dens* NEAR/4 tool*).tw.):ti,ab,kw
#25	((absorptiometr* near/4 (dpx* or dual-energ* or dual-photon* or photon*))):ti,ab,kw
#26	MeSH descriptor: [X-Rays] this term only
#27	((x-ray* or xray*)):ti,ab,kw
#28	(((grenz* or roentgen*) near/4 ray*)):ti,ab,kw
#29	((x-radiation* or xradiation*)):ti,ab,kw
#30	((DXA* or DEXA)):ti,ab,kw
#31	((FRAX or FRAXTM or Qfracture* or Q-fracture* or Cfracture* or C-Fracture*)):ti,ab,kw

#32	((fracture* near/2 risk near/2 assess* near/2 tool*)):ti,ab,kw
#33	(QCT* or pQCT* or HR-pQCT* or HRpQCT* or PCD-CT* or PCDCT* or SR-MUCT* or SRMUCT* or HRclinCT* HRclin-CT* or HR-clin-CT* or HR-clinCT*)
#34	(QUS or PEUS or P-EU or P-EUS or PEQUS)
#35	{or #21-#34}
#36	MeSH descriptor: [Tomography, X-Ray Computed] this term only
#37	((cat scan* or ct scan* or cine ct or cine-ct or tomodensitomet*)):ti,ab,kw
#38	(((computed or computer assisted or computeriz* or computeris* or electron beam* or axial*) near/4 tomograph*)):ti,ab,kw
#39	MeSH descriptor: [Four-Dimensional Computed Tomography] this term only
#40	(("4d ct" or 4dct or "4 dimensional CT" or "four dimensional CT")):ti,ab,kw
#41	MeSH descriptor: [Tomography, Spiral Computed] explode all trees
#42	(((helical or spiral) near/4 ct*)):ti,ab,kw
#43	MeSH descriptor: [Ultrasonography] explode all trees
#44	((ultrasound* or ultra-sound* or ultrason* or sonograph* or echograph* or echotomograph*)):ti,ab,kw
#45	((bindex* or echolight*)):ti,ab,kw
#46	{or #36-#45}
#47	((quantitative* or asynchronous or high-res* or highres or photon-count* or photoncount* or pulse-echo* or pulseecho* or pulsecho*)):ti,ab,kw
#48	#46 and #47
#49	#35 or #48
#50	#20 and #49
#51	((clinicaltrials or trialsearch* or trial-registry or trials-registry or clinicalstudies or trialsregister* or trialregister* or trial-number* or studyregister* or study-register* or controlled-trials-com or current-controlled-trial or AMCTR or ANZCTR or ChiCTR* or CRiS or CTIS or CTRI* or DRKS* or EU-CTR* or EUCTR* or EUDRACT* or ICTRP or IRCT* or JAPIC* or JMCTR* or JRCT or ISRCTN* or LBCTR* or NTR* or ReBec* or REPEC* or RPCEC* or SLCTR or TCTR* or UMIN*):so or (ctgov or ictrp)):an
#52	#50 not #51
#53	conference:pt
#54	#52 not #53

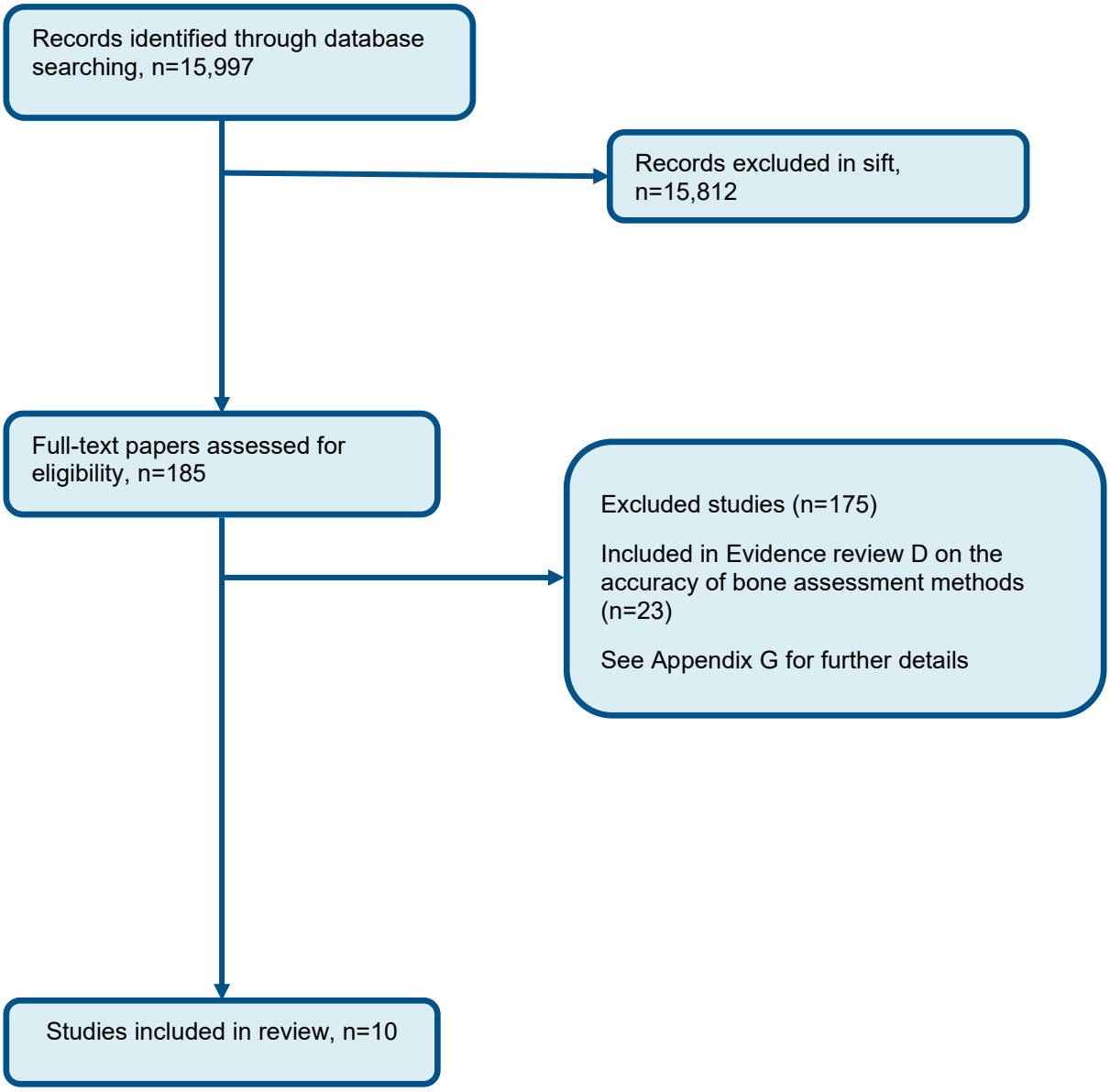
Epistemonikos search terms

1	(advanced_title_en:((osteopor* OR osteo-por* OR osteopaeni* OR osteo-paeni* OR osteopeni* OR osteo-peni*)) OR advanced_abstract_en:((osteopor* OR osteo-por* OR osteopaeni* OR osteo-paeni* OR osteopeni* OR osteo-peni*))) OR (advanced_title_en:((fragil* AND (fracture OR fractures))) OR advanced_abstract_en:((fragil* AND (fracture OR fractures)))) OR (advanced_title_en:(((low-impact* OR low-energy OR low-trauma* OR insufficien*) AND fracture*)) OR advanced_abstract_en:(((low-impact* OR low-energy OR low-trauma* OR insufficien*) AND fracture*)
2	(advanced_title_en:((advanced_title_en:((densitometr* OR BMD-test* OR densimetr*)) OR advanced_abstract_en:((densitometr* OR BMD-test* OR densimetr*))) OR (advanced_title_en:((bone AND mineral AND dens* AND test*)) OR advanced_abstract_en:((bone AND mineral AND dens* AND test*))) OR (advanced_title_en:((QCT* OR pQCT* OR HR-pQCT* OR HRpQCT* OR PCD-CT* OR PCDCT* OR SR-MUCT* OR SRMUCT* OR HRclinCT* HRclin-CT* OR HR-clin-CT* OR HR-clinCT*)) OR advanced_abstract_en:((QCT* OR pQCT* OR HR-pQCT* OR HRpQCT* OR PCD-CT* OR PCDCT* OR SR-MUCT* OR SRMUCT* OR HRclinCT* HRclin-CT* OR HR-clin-CT* OR HR-clinCT*))
3	(advanced_title_en:((QUS OR PEUS OR P-EU OR P-EUS OR PEQUS)) OR advanced_abstract_en:((QUS OR PEUS OR P-EU OR P-EUS OR PEQUS))) OR (advanced_title_en:((asynchronous OR high-res* OR highres OR photon-count* OR photoncount* OR pulse-echo* OR pulseecho* OR pulsecho*)) OR advanced_abstract_en:((asynchronous OR high-res* OR highres OR photon-count* OR photoncount* OR pulse-echo* OR pulseecho* OR pulsecho* OR risk-prediction*))
4	2 OR 3
5	1 AND 4

1

Appendix C Fragility fracture risk prediction tool evidence study selection

Figure 1: Flow chart of clinical study selection for risk prediction tools for fragility fracture review



Appendix D Fragility fracture risk prediction tools and bone assessment methods evidence

D.1.1 Akyea, 2019

Bibliographic Reference: Akyea, Ralph Kwame; McKeever, Tricia M; Gibson, Jack; Scullion, Jane E; Bolton, Charlotte E; Predicting fracture risk in patients with chronic obstructive pulmonary disease: a UK-based population-based cohort study.; BMJ open; 2019; vol. 9 (no. 4); e024951

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study included in review	NA
Study type	External validation study
Study location	UK
Study setting	Primary care
Study dates	01/2004 to 12/2015

Sources of funding	Study funded by a COPD "Open Air" research grant from Pfizer. Two authors (CEB and TMM) supported by the NIHR Nottingham BRC
Study sample	External COPD validation cohort (M/F), N=80,874 (42,799/38,075); N=72,559 (40,674/31,885) after excluding patients with READ code for osteoporosis and patients aged≥90 years
Inclusion criteria	- Inclusion criteria for validation study - Patient registered on The Health Improvement Network (THIN) primary care database - Patient aged≥40 years-old - New READ-coded COPD diagnosis 01/01/2004 to 31/12/2015 with at least 1 year record prior to this
Exclusion criteria	Exclusion criteria for validation study - History of osteoporosis or osteoporosis treatment prior to index date as determined by READ codes for osteoporosis, hip fracture, or any major osteoporotic (hip, proximal humerus, forearm, or clinically symptomatic vertebra/spine) fracture - Aged >90 years
Risk prediction tool	FRAX-UK FRAX-UK without BMD, desktop version 3.12 QFracture QFracture 2016, v.2017.0.0.0 (version 40)
Predictors	QFracture-2016 Uses same predictors as QFracture-2012 (based on QResearch, version 29), see entry for Hippisley-Cox 2012. 2016 version remodelled to account for updates to the QResearch, version 40, database. Note that the internal validation of QFracture 2016 has not been published in peer-reviewed publications but its' calibration and discrimination compared to QFracture 2012 is available at https://qfracture.org/QFracture-2016-Update-Information.pdf FRAX-UK without BMD

	- Age - Sex - Weight - Height - Prior fragility fracture - Parental history of hip fracture - Current tobacco smoking - Long-term use of oral glucocorticoids - Rheumatoid arthritis - Other causes of secondary osteoporosis - Daily alcohol consumption of three or more units daily
Risk prediction model validation	External validation of FRAX-UK without BMD and QFracture 2016 using COPD patients aged 40-90 years-old, registered on the THIN primary care database. To compared prevalence at index date and incidence, each COPD-patient matched by age, sex, and GP to up to 4 patients without COPD history to generate matched cohort and assigned same index data (N=308,999; N=264,544 after excluding patients with READ code for osteoporosis). Follow up from index date to first record of either fracture/osteoporosis, date of patient transfer out of practice area, death, or end of THIN data collection. Use of oral corticosteroids accounted for by dividing FU time into steroid-exposed (prescription date to first gap>90 days) and not exposed (from 91st day onwards) periods. Exposure effect of steroid assumed to be constant over time. Fracture outcome treated as binary variable (fracture, no fracture) and risk probabilities for FRAX and QFracture categorised according to ≥20% for MOF and ≥3% thresholds. Kaplan-Meier analysis used. - Overall discrimination assessed using AUC, with sensitivity, specificity, and positive and negative predictive values reported for ≥20% and ≥3% thresholds. - Calibration plot reported but other calibration statistics not reported.
Outcome	- 10-year risk of major osteoporotic fracture (hip, proximal humerus, forearm, or clinically symptomatic vertebra/spine) - 10-year risk of hip fracture Both outcomes confirmed with THIN database using standard READ code classification.
Duration of follow-up	Median FU, COPD patients: 5.28 years (IQR 2.6-8.3); Non-COPD patients 5.24 (IQR 2.6-8.3)

Study-level characteristics

Characteristic	Study (N = 72559)
% Female Sample size	n = 31885 ; % = 43.9
Mean age (SD) Mean (SD)	66.1 (10.7)
BMI (kg/m^2) Sample size	n = NA; % = NA
BMI - Underweight (<18.5) Sample size	n = 2730; % = 3.8
BMI - Normal (18.5-24.9) Sample size	n = 21791; % = 30
BMI - Overweight (25-29.9) Sample size	n = 21504; % = 29.6
BMI - Obese (≥30) Sample size	n = 17627; % = 24.3
BMI - No BMI	n = 8907; % = 12.3

Characteristic	Study (N = 72559)
Sample size	
Smoking status Sample size	n = NA; % = NA
Smoking status - Never smoked Sample size	n = 7062; % = 9.7
Smoking status - Ex-smoker Sample size	n = 33810; % = 46.6
Smoking status - Current smoker Sample size	n = 29949; % = 41.3
Smoking status - Unknown Sample size	n = 1738; % = 2.4
Fall history Prior to or at diagnosis. Data is for N=80,874. Sample size	n = NA; % = NA
Fall history - Personal history (N=80,874 population before excluded people with osteoporosis or over 90 years)	n = 8969; % = 11.1
Fall history - Parental history of fall/osteoporosis	n = 96; % = 0.1

Characteristic	Study (N = 72559)
(N=80,874 population before excluded people with osteoporosis or over 90 years)	
Corticosteroid use Data for N=80874. Sample size	n = NA; % = NA
Corticosteroid use - Inhaled corticosteroid use Sample size	n = 47574; % = 58.8
Corticosteroid use - Oral corticosteroid use Sample size	n = 33618; % = 41.6
MRC Dyspnoea Scale (Lower scores are better) 1 year either side of diagnosis Sample size	n = NA; % = NA
MRC Dyspnoea Scale - Score=1 Sample size	n = 9499; % = 11.8
MRC Dyspnoea Scale - Score=2 Sample size	n = 19466; % = 24.1
MRC Dyspnoea Scale - Score=3 Sample size	n = 10488; % = 13
MRC Dyspnoea Scale - Score 4 & 5	n = 5237; % = 6.5

Characteristic	Study (N = 72559)
Sample size	
MRC Dyspnoea Scale - No record Sample size	n = 36184; % = 44.7
Charlson Comorbidity Index Score Lower scores are better Sample size	n = NA; % = NA
Charlson Comorbidity Index Score - Score=0 Sample size	n = 0; % = 0
Charlson Comorbidity Index Score - Score=1 Sample size	n = 38573; % = 53.2
Charlson Comorbidity Index Score - Score=2 Sample size	n = 11953; % = 16.5
Charlson Comorbidity Index Score - Score=3 Sample size	n = 11110; % = 15.3
Charlson Comorbidity Index Score - Score≥4 Sample size	n = 10923; % = 15.1

Outcomes

FRAX-UK outcomes

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 67954, N1 = 4605	Hip fracture vs No hip fracture, N2 = 71115, N1 = 1444
AUC Custom value	0.714 (95%CI 0.706-0.722)	0.761 (95%CI 0.749-0.772)
Sensitivity 20% fracture risk threshold for MOF; 3% fracture risk threshold for HF Custom value	25.4% (95%CI 22.7-28.1)	78.1% (95%CI 75.6-80.7)
Specificity 20% fracture risk threshold for MOF; 3% fracture risk threshold for HF Custom value	92.6% (95%CI 91.0-94.2)	60.8% (95%CI 57.8-63.8)
Positive predictive value 20% fracture risk threshold for MOF; 3% fracture risk threshold for HF Custom value	18.8% (95%CI 16.4-21.1)	3.9% (95%CI 2.7-5.1)
Negative predictive value 20% fracture risk threshold for MOF; 3% fracture risk threshold for HF Custom value	94.8% (95%CI 93.4-96.2)	99.3% (95%CI 98.8-99.8)

Sensitivity - Polarity - Higher values are better

Specificity - Polarity - Higher values are better

Positive predictive value - Polarity - Higher values are better

Negative predictive value - Polarity - Higher values are better

n1=fractured group, n2=non-fractured group

QFracture (2016 version) outcomes

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 67954, N1 = 4605	Hip fracture vs No hip fracture, N2 = 71115, N1 = 1444
AUC Custom value	0.614 (95%CI 0.605-0.623)	0.761 (95%CI 0.749-0.772)
Sensitivity 20% fracture risk threshold for MOF; 3% fracture risk threshold for HF Custom value	25.2% (95%CI 22.5-27.9)	82.1% (95%CI 79.7-84.5)
Specificity 20% fracture risk threshold for MOF; 3% fracture risk threshold for HF Custom value	87.7% (95%CI 85.7-89.7)	55.2% (95%CI 52.1-58.3)
Positive predictive value 20% fracture risk threshold for MOF; 3% fracture risk threshold for HF Custom value	12.2% (95%CI 10.2-14.2)	3.6% (95%CI 2.5-4.8)

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 67954, N1 = 4605	Hip fracture vs No hip fracture, N2 = 71115, N1 = 1444
Negative predictive value 20% fracture risk threshold for MOF; 3% fracture risk threshold for HF Custom value	94.5% (95%CI 93.1-95.9)	99.3% (95%CI 98.8-99.8)

AUC - Polarity - Higher values are better

Sensitivity - Polarity - Higher values are better

Specificity - Polarity - Higher values are better

Positive predictive value - Polarity - Higher values are better

Negative predictive value - Polarity - Higher values are better

n1=fractured group, n2=non-fractured group

Critical appraisal – PROBAST tool for risk prediction models

Overall Risk of bias and Applicability	Risk of bias	High *(FRAX-UK/QFracture 2016: High RoB on participants (excludes people with history of osteoporosis or osteoporosis treatment at index date), predictors (presence of various diseases/conditions likely to be variable), outcomes (use of GP database to assess fracture occurrence), and analysis (Kaplan-Meier analysis used, does not account of competing mortality risk; calibration plot/statistics not reported) domains)*
Overall Risk of bias and Applicability	Directness	Low

D.1.2 Collins, 2011

Bibliographic Reference Collins, Gary S; Mallett, Susan; Altman, Douglas G; Predicting risk of osteoporotic and hip fracture in the United Kingdom: prospective independent and external validation of QFractureScores.; BMJ (Clinical research ed.); 2011; vol. 342; d3651

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study included in review	Original QFracture development and external development study Hippisley-Cox, J., & Coupland, C. (2009). Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores. *Bmj*, *339*.
Study type	External validation study
Study location	UK
Study setting	Primary care
Study dates	06/1994 to 06/2008
Sources of funding	No specific grant was received from any funding agency in public, commercial, and not-for-profit sectors
Study sample	External validation cohort (M/F), N=2,244,636 (1,108,219/1,136,417) from 364 UK GPs
Inclusion criteria	- Patient registered on THIN database (GPs using INPS Vision computer system [In Practice Systems, London]) with minimum 1-year complete data in their medical record

	- Aged between 30-85 years - No previously recorded hip, distal radius, or vertebral fracture - Permanent UK resident - No uninterrupted GP registration period
Exclusion criteria	NA
Population subgroups	- Men - Women
Risk prediction tool	QFracture 2009 version
Predictors	QFracture 2009 See list of predictors for Hippisley-Cox 2009.
Risk prediction model validation	External validation study of QFracture, 2009 version, using patients of 364 UK GPs registered on THIN database. Observed 10-year fracture risk calculated for every patient in THIN cohort using Kaplan-Meier method. Missing data for smoking status, number of cigarettes smoked, alcohol consumption, and BMI replaced using multiple imputation (MI) using all predictors and outcome variable (major osteoporotic fracture or hip fracture as appropriate). Multiple copies of data created with missing values imputed with sensible values randomly selected from predicted distribution. Five imputed datasets generated, and results combined using Rubin's rules to allow uncertainty of imputed values to be incorporated. Kaplan-Meier analysis used. - Calibration assessed by: observed/predicted 10-year fracture risk for each 10th (decile) of risk and for each 7-year age band, calibration plot. - Overall fit assessed by: R^2, Brier score. - Discrimination assessed by: AUC and D statistic.
Outcome	- 10-year risk of (incident) major osteoporotic (hip, distal radius, or vertebral) fracture - 10-year risk of (incident) hip fracture

	Both outcomes confirmed by THIN database codes.
Duration of follow-up	Median FU for major osteoporotic fracture=5.98 years (IQR 2.61-8.50) Median FU for hip fracture=6.03 years (2.62-8.50)

Study-level characteristics

Characteristic	Study (N = 2244636)
% Female Sample size	n = 1136417; % = 50.6

Arm-level characteristics

Characteristic	**Male (N = 1108219)**	**Female (N = 1136417)**
Mean age (SD) Median (IQR)	47 (37 to 59)	48 (37 to 62)
BMI BMI recorded for 25.6% of men and 82.4% of women **Mean (SD)**	26.63 (4.1)	26.15 (5)
Alcohol intake Sample size	n = NA; % = NA	n = NA; % = NA
Alcohol intake - Not recorded	n = 672709; % = 60.7	n = 511776; % = 45

Characteristic	Male (N = 1108219)	Female (N = 1136417)
Sample size		
Alcohol intake - None Sample size	n = 132872; % = 12	n = 243624; % = 21.4
Alcohol intake - <1 unit/day Sample size	n = 168374; % = 15.2	n = 288754; % = 25.4
Alcohol intake - 1-2 units/day Sample size	n = 72962; % = 6.6	n = 71616; % = 6.3
Alcohol intake - 3-6 units/day Sample size	n = 48270; % = 4.4	n = 17911; % = 1.6
Alcohol intake - 7-9 units/day Sample size	n = 7986; % = 0.7	n = 1550; % = 0.1
Alcohol intake - >9 units/day Sample size	n = 5046; % = 0.5	n = 1178; % = 0.1
Smoking status Sample size	n = NA; % = NA	n = NA; % = NA
Smoking status - Not recorded Sample size	n = 119754; % = 10.8	n = 69470; % = 6.1

Characteristic	Male (N = 1108219)	Female (N = 1136417)
Smoking status - Non-smoker Sample size	n = 401760; % = 36.3	n = 530062; % = 46.6
Smoking status - Former smoker Sample size	n = 158600; % = 14.3	n = 125816; % = 11.1
Smoking status - Current light smoker <10 cigarettes/day Sample size	n = 68077; % = 6.1	n = 70741; % = 6.2
Smoking status - Current moderate smoker 10-19 cigarettes/day Sample size	n = 104844; % = 9.5	n = 109052; % = 9.6
Smoking status - Current heavy smoker ≥20 cigarettes/day Sample size	n = 117567; % = 10.6	n = 77828; % = 6.9
Fall history Sample size	n = 14911; % = 1.4	n = 29106; % = 2.6
Cardiovascular disease Sample size	n = 76585; % = 6.9	n = 54520; % = 4.8
Chronic liver disease	n = 2586; % = 0.2	n = 1892; % = 0.2

Characteristic	Male (N = 1108219)	Female (N = 1136417)
Sample size		
Endocrine disorders Sample size	n = 2124; % = 0.2	n = 9665; % = 0.9
Gastrointestinal malabsorption | Sample size	n = 5047; % = 0.5	n = 6388; % = 0.6
Rheumatoid arthritis or systemic lupus erythematosus Rheumatoid arthritis only Sample size	n = 5260; % = 0.5	n = 12340; % = 1.1
Type 2 diabetes No of events	n = 35157; % = 3.2	n = 28039; % = 2.5
Antidepressant use Current tricyclic antidepressant use Sample size	n = 23048; % = 2.1	n = 59803; % = 5.3
Corticosteroid use Current use Sample size	n = 23686; % = 2.1	n = 36752; % = 3.2
Menopausal symptoms Sample size	n = NA; % = NA	n = 58507; % = 5.2

Outcomes

QFracture (2009 version) outcomes in men

Outcome	Osteoporotic fracture in men vs No osteoporotic fracture in men,N2 = 1102066, N1 = 6153	Hip fracture in men vs No hip fracture in men, , N2 = 1105196, N1 = 3023
R^2 (R-squared) (%) Custom value	37.99 (95%CI 36.64-39.35)	60.42 (95%CI 59.22-61.63)
Brier score Values range from 0 (perfect accuracy) to 1 (inaccurate) Custom value	0.010 (95%CI 0.008-0.012)	0.005 (95%CI 0.003-0.007)
AUC Custom value	0.739 (95%CI 0.733-0.745)	0.855 (95%CI 0.848-0.862)
D-statistic Custom value	1.60 (95%CI 1.56-1.65)	2.53 (95%CI 2.46-2.59)

R^2 (R-squared) - Polarity - Higher values are better

Brier score - Polarity - Lower values are better

AUC - Polarity - Higher values are better

D-statistic - Polarity - Higher values are better

n1=fractured group, n2=non-fractured group

QFracture (2009 version) outcomes in women

Outcome	Osteoporotic fracture in women vs No osteoporotic fracture in women, N2 = 1117362, N1 = 19055	Hip fracture in women vs No hip fracture in women, N2 = 1127252, N1 = 9165
R^2 (R-squared) (%) Custom value	49.24 (95%CI 48.64-49.85)	62.82 (95%CI 62.22-63.43)
Brier score Values range from 0 (perfect accuracy) to 1 (inaccurate) Custom value	0.027 (95%CI 0.025-0.029)	0.013 (95%CI 0.012-0.015)
AUC Custom value	0.816 (95%CI 0.813-0.819)	0.890 (95%CI 0.887-0.893)
D-statistic Custom value	2.02 (95%CI 1.99-2.04)	2.66 (95%CI 2.63-2.70)

R^2 (R-squared) - Polarity - Higher values are better

Brier score - Polarity - Lower values are better

AUC - Polarity - Higher values are better

D-statistic - Polarity - Higher values are better

n1=fractured group, n2=non-fractured group

Critical appraisal – PROBAST tool for risk prediction models

Overall Risk of bias and Applicability	Risk of bias	High *(QFracture 2009: High RoB for participants (excludes people with history of fracture and those without recorded Townsend score), predictors (diagnosis of disease/condition likely to have been variable across*

		participants), outcome (use of GP records likely to underestimate fracture occurrence), and analysis (Kaplan-Meier analysis used, does not account for competing mortality risk) domains)
Overall Risk of bias and Applicability	Directness	Partially applicable *(Includes participants not at suspected risk or not at risk of fragility fracture; includes women < 65 and men < 75 years)*

D.1.3 Green, 2024

Bibliographic Reference Green, Olivia; Raju, Suneil A; Shiha, Mohamed G; Nandi, Nicoletta; Bayley, Martin; McCloskey, Eugene; Sanders, David S; Clinical utility of the fracture risk assessment tool (FRAX) in biopsy-confirmed coeliac disease.; Scandinavian journal of gastroenterology; 2024; vol. 59 (no. 9); 1049-1054

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study included in review	NA
Study type	External validation study Retrospective cohort study
Study location	Sheffield, UK

Study setting	Tertiary care (single-centre tertiary hospital)
Study dates	2001 to 2015
Sources of funding	Reports no funding received
Study sample	External validation cohort (M/F), N=593 (187/406) adults with biopsy-proved coeliac disease
Inclusion criteria	- Aged≥18 years - Biopsy-proven diagnosis of coeliac disease (Marsh≥3a) between 2001 and 2015 at tertiary hospital in Sheffield, UK - Had DXA scan within 1 year of diagnosis Note: For patient with discordant histology in D1 and D2, overall grade defined as the most severe grade for individual. Diagnosis date taken to be date of first endoscopy where positive biopsies obtained
Population subgroups	NA
Risk prediction tool	FRAX-UK FRAX Desktop Multi-Patient Entry software with and without FN BMD, with National Osteoporosis Guidelines Group (NOGG) guidelines thresholds
Predictors	<u>FRAX-UK</u> - Age - Sex - Weight - Height - Prior fragility fracture - Parental history of hip fracture - Current tobacco smoking - Long-term use of oral glucocorticoids

	- Rheumatoid arthritis - Other causes of secondary osteoporosis - Daily alcohol consumption of three or more units daily
Risk prediction model validation	Patients identified from coeliac disease database and another database containing participants' DXA scan data. Femoral neck (FN) BMD obtained from DXA scanner and reported as T-score. Data also collected on serology at diagnosis; IgA-tTG antibody levels measured; gluten exposure determined by review of clinical notes and assessment of each patient's ongoing serology. All patients reviewed by dietician for minimum of 1-year post-diagnosis. Self-reported patient questionnaire completed on attendance for DXA scan. FRAX-UK scores calculated using FRAX software with and without FN BMD. Following National Osteoporosis Guidelines Group (NOGG) guidelines, patient categorized as 'high' or 'low' risk of major osteoporotic fracture (MOF) according to their FRAX score (with and without BMD) and FRAX-threshold for their age (for example, the FRAX score for high risk of MOF for a 50 year-old is 7.3, whilst for a 70-year old is 20.3). Patients aged <50 years were assigned same risk threshold as those aged 50, whilst patients >70 years were assigned same risk threshold as those aged 70 years. Fracture outcomes identified via search of individual patient records on virtual healthcare platform (including site, nature of fracture, date of fracture) and clinical review. - Discrimination at NOGG thresholds assessed by sensitivity, specificity, and positive and negative predictive values. Calibration statistics not reported.
Outcome	- 10-year risk of major osteoporotic fracture Outcome confirmed by patient records and clinical review.
Duration of follow-up	Median FU=10.5 years (IQR 9.0-13.4)

Characteristics

Study-level characteristics

Characteristic	**Study (N = 593)**
% Female	n = 406; % = 68.5
Sample size	

Characteristic	Study (N = 593)
Mean age (SD) (years) Median (IQR)	45 (31.5 to 57.6)
BMI (kg/m^2) Median (IQR)	24.5 (21.6 to 28.5)
Alcohol intake >3 units/day Sample size	n = 42; % = 7.1
Smoking status Current smoker Sample size	n = 102; % = 17.2
Previous fracture Sample size	n = 50; % = 8.4
Parental history of osteoporosis History of parental hip fracture Sample size	n = 22; % = 3.7
Rheumatoid arthritis or systemic lupus erythematosus Rheumatoid arthritis only Sample size	n = 14; % = 2.4

Characteristic	Study (N = 593)
Corticosteroid use Oral glucocorticoid only Sample size	n = 15; % = 2.6
Vitamin D deficiency Sample size	n = 76; % = 12.8
Calcium deficiency Sample size	n = 65; % = 11
Marsh histological grade Sample size	n = NA; % = NA
Marsh histological grade - 3a Sample size	n = 158; % = 26.6
Marsh histological grade - 3b Sample size	n = 186; % = 31.4
Marsh histological grade - 3c Sample size	n = 249; % = 42
Ongoing gluten exposure Sample size	n = 109; % = 18.4

Outcomes

FRAX-UK outcomes

Outcome	**Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 538, N1 = 55**
Sensitivity NOGG age-specific thresholds Custom value	NA
Sensitivity - FRAX without BMD Custom value	22.0 % (95%CI 12.0-35.0)
Sensitivity - FRAX with BMD Custom value	15% (95%CI 6.0-27.0)
Specificity NOGG age-specific thresholds Custom value	NA
Specificity - FRAX without BMD Custom value	91.0% (95%CI 89.0-94.0)
Specificity - FRAX with BMD Custom value	92.0% (95%CI 89.0-94.0)
Positive predictive value NOGG age-specific thresholds Custom value	NA

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 538, N1 = 55
Positive predictive value - FRAX without BMD Custom value	16.3% (95%CI 8.7-27.6)
Positive predictive value - FRAX with BMD Custom value	11.4% (95%CI 4.9-22.6)
Negative predictive value NOGG age-specific thresholds Custom value	NA
Negative predictive value - FRAX without BMD Custom value	93.5% (95%CI 91.1-95.3)
Negative predictive value - FRAX with BMD Custom value	93.4% (95%CI 91.0-95.2)

Sensitivity - Polarity - Higher values are better

Specificity - Polarity - Higher values are better

Positive predictive value - Polarity - Higher values are better

Negative predictive value - Polarity - Higher values are better

n1=fractured group, n2=non-fractured group. Discrimination outcomes are reported using the NOGG age-specific thresholds.

Critical appraisal – PROBAST tool for risk prediction models

Overall Risk of bias and Applicability	Risk of bias	High *(FRAX-UK: High RoB for participant (excludes people at suspected risk of coeliac disease and therefore fragility fracture), predictor (rheumatoid arthritis diagnosis likely to be variable) and analysis (<100 participants with fracture; no info on missing data strategy; calibration not reported).)*
Overall Risk of bias and Applicability	Directness	Low

D.1.4 Hippisley-Cox, 2009

Bibliographic Reference Hippisley-Cox, Julia; Coupland, Carol; Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFracture Scores.; BMJ (Clinical research ed.); 2009; vol. 339; b4229

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study included in review	NA
Study type	Internal validation study

	QFracture, 2009 version External validation study FRAX-UK
Study location	UK
Study setting	Primary care
Study dates	01/1993 to 06/2008
Sources of funding	Study funded by David Stables (Medical director of EMIS) as part of a larger study examining risks and benefits of HRT.
Study sample	Derivation cohort (M/F), N=2,391,576 (1,187,354/1,204,222) from 357 GP practices; N=2,357,895 (1,174,232/1,183,663) after excluding patients with previous fracture before start of study Validation cohort (M/F), N=1,294,732 (640,943/653,789) from 178 GP practices; N=1,275,917 (633,764/642,153) after excluding patients with previous fracture before start of study
Inclusion criteria	- Patients registered on QResearch database, version 20 (574 registered GPs that use Egton Medical Information System [EMIS] computer system) - Registered GPs used EMIS system for at least 1 year - Patients aged 30-85 at study entry date - Patients registered with GP from 01/01/1993 to 30/06/2008
Exclusion criteria	- Patients with previous recorded hip, distal radius, or vertebral fracture - Temporary residents - Patients with interrupted periods of registration with GP - Patients without valid Townsend deprivation score related to their postcode
Population subgroups	- Men - Women

Risk prediction tool	FRAX-UK FRAX without BMD, hip fracture QFracture 2009 version, major osteoporotic fracture, hip fracture
Predictors	**<u>QFracture 2009</u>** *Men (Major osteoporotic fracture, hip fracture)* - Age - BMI - Smoking status - Alcohol use - Fall history - Asthma - Cardiovascular disease - Liver disease - Rheumatoid arthritis - Type 2 diabetes - Current corticosteroid use - Tricyclic anti-depressants use - Women (Major osteoporotic fracture) - Smoking status - Alcohol use - Fall history - Parental history of osteoporosis - Menopausal symptoms - Asthma - Cardiovascular disease

- Chronic liver disease
- Gastrointestinal malabsorption
- Other endocrine disorders
- Rheumatoid arthritis
- Type 2 diabetes
- Current corticosteroid use
- Hormone replacement therapy use
- Tricyclic anti-depressants use

Women (Hip fracture)

- Age
- BMI
- Smoking status
- Alcohol use
- Fall history
- Parental history of osteoporosis
- Asthma
- Cardiovascular disease
- Chronic liver disease
- Gastrointestinal malabsorption
- Other endocrine disorders
- Rheumatoid arthritis
- Type 2 diabetes
- Current corticosteroid use
- Hormone replacement therapy use
- Tricyclic anti-depressants use

FRAX-UK - Hip fracture

- Age
- Sex
- Height

	- Weight - Smoking status - Alcohol use - Fracture history - Parental history of hip fracture - Secondary osteoporosis - Glucocorticoid treatment - Rheumatoid arthritis
Risk prediction model validation	Internal validation of QFracture, 2009 Development (derivation) of QFracture conducted on two-thirds of registered GPs (357 practices) by random allocation (simple random sampling utility, STATA) with remaining one-third (178 practices) comprising the validation dataset. Missing values for alcohol use, smoking status, and BMI replaced using multiple imputation. Patients with no recorded values for diagnosis, prescription, or family history were assumed not to be exposed. - Calibration assessed by: Mean predicted v observed 10-year fracture risk compared by deciles using Kaplan-Meier method - Overall fit assessed by: R^2 statistic - Discrimination assessed by: AUC and D statistic - Comparison to FRAX-UK External validation of FRAX-UK without BMD for hip fracture assessed using same performance measures as above, and compared to QFracture, with cohort limited to people aged 40-85 years. Previous fractures counted as negative given QFracture restricted to patients without previous fracture. Missing values for alcohol use, smoking status and BMI treated in same way as above. Variables for each patient entered twice using automated software to test reproducibility of FRAX-UK scores.
Outcome	- 10-year risk of major osteoporotic fracture (Hip, vertebral, or distal radius) - 10-year risk of hip fracture Both outcomes confirmed by READ codes on QResearch primary care database.
Duration of follow-up	Follow up for derivation cohort reported as 7,898,208 person years for women, and 8,049,306 person years for men. Follow up not reported for validation cohort.

Characteristics

Study-level characteristics

Characteristic	Study (N = 1275917)
% Female Sample size	n = 642153; % = 50.3

Arm-level characteristics

Characteristic	Male (N = 633764)	Female (N = 642153)
Mean age (SD) Median (IQR)	46 (37 to 69)	49 (37 to 63)
BMI (kg/m^2) BMI recorded in ~68% of male and 75% of female participants Mean (SD)	26.41 (4.02)	25.82 (4.85)
Alcohol intake Sample size	n = NA; % = NA	n = NA; % = NA
Alcohol intake - Recorded Sample size	n = 391290; % = 61.74	n = 435452; % = 67.81
Alcohol intake - Non-drinker	n = 74718; % = 11.79	n = 148646; % = 23.15

Characteristic	Male (N = 633764)	Female (N = 642153)
Sample size		
Alcohol intake - <1 unit/day	n = 120989; % = 19.09	n = 185570; % = 28.9
Sample size		
Alcohol intake - 1-2 units/day	n = 130813; % = 20.64	n = 89435; % = 13.93
Sample size		
Alcohol intake - 3-6 units/day	n = 54239; % = 8.56	n = 10610; % = 1.65
Sample size		
Alcohol intake - 7-9 units	n = 6005; % = 0.95	n = 618; % = 0.1
Sample size		
Alcohol intake - >9 units/day	n = 4567; % = 0.72	n = 616; % = 0.1
Sample size		
Smoking status	n = NA; % = NA	n = NA; % = NA
Sample size		
Smoking status - Recorded	n = 502739; % = 79.33	n = 547531; % = 85.26
Sample size		
Smoking status - Non-smoker	n = 250715; % = 39.56	n = 340811; % = 53.07
Sample size		

Characteristic	Male (N = 633764)	Female (N = 642153)
Smoking status - Ex-smoker Sample size	n = 95004; % = 14.99	n = 75629; % = 11.78
Smoking status - Current light smoker Sample size	n = 38173; % = 6.02	n = 29288; % = 4.56
Smoking status - Current moderate smoker Sample size	n = 76908; % = 12.14	n = 71638; % = 11.16
Smoking status - Current heavy smoker Sample size	n = 41939; % = 6.62	n = 30165; % = 4.7
Parental history of osteoporosis Sample size	n = 128; % = 0.02	n = 2180; % = 0.34
Fall history Sample size	n = 3036; % = 0.48	n = 2180; % = 0.34
Cardiovascular disease Sample size	n = 33542; % = 5.29	n = 23375; % = 3.64
Chronic liver disease Sample size	n = 1205; % = 0.19	n = 809; % = 0.13

Characteristic	Male (N = 633764)	Female (N = 642153)
Chronic obstructive airways disease or asthma Asthma only Sample size	n = 29992; % = 4.57	n = 35081; % = 5.46
Endocrine disorders Sample size	n = 1044; % = 0.16	n = 5039; % = 0.78
Gastrointestinal malabsorption Sample size	n = 2595; % = 0.41	n = 3346; % = 0.52
Rheumatoid arthritis or systemic lupus erythematosus Rheumatoid arthritis only Sample size	n = 2114; % = 0.33	n = 5013
Type 2 diabetes Sample size	n = 14257; % = 2.25	n = 11919; % = 1.86
Antidepressant use Tricyclic antidepressant use Sample size	n = 7354; % = 1.16	n = 23729; % = 3.7
Corticosteroid use Current use Sample size	n = 5792; % = 0.91	n = 10509; % = 1.64
Menopausal symptoms	n = NA; % = NA	n = 11830; % = 1.84

Characteristic	Male (N = 633764)	Female (N = 642153)
Sample size		

Note that baseline characteristics are for validation cohort

Outcomes

FRAX-UK outcomes

Outcome	**Hip fracture in men vs No hip fracture in men, N2 = 632026, N1 = 1738**	**Hip fracture in women vs No hip fracture in women, N2 = 636729, N1 = 5424**
R^2 (R-squared) (%) Custom value	54.07 (95%CI 52.1-53.65)	54.83 (95%CI 54.43-55.12)
O:E Ratio O:E ratio calculated using equation in Table 2, Debray 2018; Custom value	0.741 (95%CI 0.673-0.808)	0.868 (95%CI 0.815-0.921)
AUC 95% CI calculated from reported total number of patients, number of incident fractures, and number of patients not experiencing incident fractures using equations in Table 2, Debray 2018. Custom value	0.817 (95%CI 0.807-0.827)	0.845 (95%CI 0.840-0.850)
D-statistic Custom value	2.22 (95%CI 2.14-2.3)	2.26 (95%CI 2.21-2.3)

R^2 (R-squared) - Polarity - Higher values are better

AUC - Polarity - Higher values are better

D-statistic - Polarity - Higher values are better

n1=fractured group, n2=non-fractured group

QFracture (2009 version) outcomes in men

Outcome	Major osteoporotic fracture in men vs No major osteoporotic fracture in men, N2 = 629245, N1 = 4519	Hip fracture in men vs No hip fracture in men N2 = 632026, N1 = 1738
R^2 (R-squared) (%) Custom value	30.02 (95%CI 22.21-37.84)	63.19 (95%CI 60.81-65.57)
O:E Ratio O:E ratio calculated using equations in Table 2, Debray 2018 Custom value	0.984 (95%CI 0.953-1.014)	0.879 (95%CI 0.767-0.992)
O:E Ratio - People aged 40-85 years Custom value	*empty data*	0.906 (95%CI 0.817-0.994)
AUC Custom value	0.688 (95%CI 0.684-0.692)	0.856 (95%CI 0.851-0.860)
D-statistic Custom value	1.34 (95%CI 1.09-1.59)	2.68 (95%CI 2.55-2.82)

R^2 (R-squared) - Polarity - Higher values are better

AUC - Polarity - Higher values are better

D-statistic - Polarity - Higher values are better

QFracture (2009 version) outcomes in women

Outcome	Major osteoporotic fracture in women vs No major osteoporotic fracture in women, N2 = 628201, N1 = 13952	Hip fracture in women vs No hip fracture in women, N2 = 636729, N1 = 5424
R^2 (R-squared) (%) Custom value	44.87 (95%CI 43.07-46.67)	63.94 (95%CI 62.12-65.76)
O:E Ratio O:E ratio calculated using equations in Table 2, Debray 2018 Custom value	0.999 (95%CI 0.975-1.023)	0.968 (95%CI 0.852-1.084)
O:E Ratio - People aged 40-85 years Custom value	NR	0.982 (95%CI 0.900-1.065)
AUC Custom value	0.788 (95%CI 0.786-0.790)	0.890 (95%CI 0.889-0.892)
D-statistic Custom value	1.85 (95%CI 1.78-1.91)	2.73 (95%CI 2.62-2.83)

R^2 (R-squared) - Polarity - Higher values are better

AUC - Polarity - Higher values are better

D-statistic - Polarity - Higher values are better

n1=fractured group, n2=non-fractured group

Critical appraisal – PROBAST tool for risk prediction models

Overall Risk of bias and Applicability	Risk of bias	High *(FRAX-UK/QFracture 2009: High risk of bias for participants (excludes previous fracture), predictors (diagnosis of disease/condition likely to have been variable across participants), outcome (use of GP records for fracture occurrence; excludes proximal humerus fractures), and analysis (Kaplan-Meier analysis used, does not account for competing risks) domains.)*
Overall Risk of bias and Applicability	Directness	Partially applicable *(includes participants not at suspected risk or not at risk of fragility fracture; women < 65 and men < 75 years)*

D.1.5 Hippisley-Cox, 2012

Bibliographic Reference	Hippisley-Cox, Julia; Coupland, Carol; Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study.; BMJ (Clinical research ed.); 2012; vol. 344; e3427

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study	NA

included in review	
Study type	Internal validation study Internal validation of QFracture-2012 External validation study External validation of QFracture 2009
Study location	UK
Study setting	Primary care
Study dates	01/1993 to 10/2011
Sources of funding	No external source of funding
Study sample	Internal validation cohort (M/F), N=1,583,373 (778,810/804,563) from 207 UK GPs
Inclusion criteria	- Patients registered on QResearch database, version 32 (627 registered GPs that use Egton Medical Information System [EMIS] computer system) - Registered GPs used EMIS system for at least 1 year - Patients aged 30-100 at study entry date - Patients registered with GP from 01/01/1993 to 01/10/2011 - Patients with previous recorded fracture were eligible for inclusion in derivation and validation dataset
Population subgroups	- Men - Women
Risk prediction tool	QFracture - External validation of 2009 version - Internal validation of 2012 version

Predictors	**QFracture 2012** *Men and Women (Major osteoporotic fracture, Hip fracture)* - Age - BMI - Ethnic origin - Smoking status - Alcohol use - Fall history - Parental history of osteoporosis (not significant predictor for women's hip fracture risk) - Fracture history - Any cancer - Cardiovascular disease - Chronic liver disease - Chronic renal disease - Chronic obstructive airways disease or asthma - Dementia - Endocrine disorders - Epilepsy diagnosis or treatment - Gastrointestinal malabsorption (not significant predictor for men and women's hip fracture) - Parkinson's disease - Rheumatoid arthritis or systemic lupus erythematosus - Type 1 diabetes - Type 2 diabetes - Anti-depressants use - Current corticosteroid use - Hormone replacement therapy use - Care home residency (not significant predictor for women)

Risk prediction model validation	QFracture, 2012 version Development (derivation) of QFracture conducted on two-thirds of registered GPs by random allocation (simple random sampling utility, STATA) with remaining one-third of GPs comprising the validation dataset. Final development models fit using Rubin's rules to allow uncertainty due to missing data to be incorporated. Regression coefficients (log of hazard ratios) from final models used as weights and combined with baseline survivor functions for (major osteoporotic, hip) fracture to derive 10-year fracture risk equation. Missing values for alcohol use, smoking status, and BMI replaced using multiple imputation. No sample size calculation conducted because all available data from QResearch database was used. - Calibration assessed by: Mean predicted v observed 10-year fracture risk using Kaplan-Meier method by every tenth [decile] of risk). Kaplan-Meier analysis used. - Overall fit assessed by: R^2 statistic - Discrimination assessed by: AUC, D statistic - Reclassification assessed by: Net reclassification of cases by QFracture 2012 version compared to QFracture 2009 version.
Outcome	- 10-year risk of major osteoporotic (hip, vertebral, proximal humerus, distal radius) fracture - 10-year risk of hip fracture Both outcomes confirmed by GP computer record or linked death record.
Duration of follow-up	Reports follow up of 11,732,106 person years (N=1,583,373), FU≈7.41 years

Characteristics

Study-level characteristics

Characteristic	Study (N = 1583373)
% Female	n = 804863; % = 50.8

Characteristic	Study (N = 1583373)
Sample size	
Mean age (SD) Mean (SD)	50 (16)
Ethnicity Sample size	n = NA; % = NA
Ethnicity - Ethnic origin recorded Sample size	n = 727888; % = 46
Ethnicity - White or not recorded Sample size	n = 1493455; % = 94.3
Ethnicity - Bangladeshi Sample size	n = 4191; % = 0.3
Ethnicity - Indian Sample size	n = 17670; % = 1.1
Ethnicity - Pakistani Sample size	n = 6489; % = 0.4
Ethnicity - Other Asian Sample size	n = 10779; % = 0.7

Characteristic	Study (N = 1583373)
Ethnicity - Black African Sample size	n = 17367; % = 1.1
Ethnicity - Caribbean Sample size	n = 10144; % = 0.6
Ethnicity - Chinese Sample size	n = 5206; % = 0.3
Ethnicity - Other Sample size	n = 18072; % = 1.1
BMI (kg/m^2) BMI recorded for 76.6% of adults Mean (SD)	26.1 (4.6)
Alcohol intake Sample size	n = NA; % = NA
Alcohol intake - Not reported Sample size	n = 461740; % = 29.2
Alcohol intake - None Sample size	n = 330695; % = 20.9

Characteristic	Study (N = 1583373)
Alcohol intake - <1 unit/day Sample size	n = 402847; % = 25.4
Alcohol intake - 1-2 units/day Sample size	n = 287441; % = 18.2
Alcohol intake - 3-6 units/day Sample size	n = 84478; % = 5.3
Alcohol intake - 7-9 units/day Sample size	n = 7429; % = 0.5
Smoking status Sample size	n = NA; % = NA
Smoking status - Not recorded Sample size	n = 193038; % = 12.2
Smoking status - Non-smoker Sample size	n = 773198; % = 48.8
Smoking status - Past smoker Sample size	n = 257087; % = 16.2
Smoking status - Current light smoker	n = 94400; % = 6

Characteristic	Study (N = 1583373)
Sample size	
Smoking status - Current moderate smoker Sample size	n = 113757; % = 7.2
Smoking status - Current heavy smoker Sample size	n = 86787; % = 5.5
Previous fracture Sample size	n = 27907; % = 1.8
Parental history of osteoporosis Sample size	n = 4227; % = 0.3
Fall history Sample size	n = 17382; % = 1.1
Cardiovascular disease Sample size	n = 77824; % = 4.9
Cancer Sample size	n = 28203; % = 1.8
Chronic liver disease Sample size	n = 3216; % = 0.2

Characteristic	Study (N = 1583373)
Chronic renal disease Sample size	n = 3413; % = 0.2
Chronic obstructive airways disease or asthma Sample size	n = 113175; % = 7.1
Dementia Sample size	n = 7791; % = 0.5
Endocrine disorders Sample size	n = 7882; % = 0.5
Epilepsy Diagnosis or treatment **Sample size**	n = 26271; % = 1.7
Gastrointestinal malabsorption Sample size	n = 8026; % = 0.5
Parkinson's disease Sample size	n = 3650; % = 0.2
Rheumatoid arthritis or systemic lupus erythematosus Sample size	n = 10091; % = 0.6

Characteristic	Study (N = 1583373)
Type 1 diabetes Sample size	n = 4322; % = 0.3
Type 2 diabetes Sample size	n = 43437; % = 2.7
Antidepressant use Sample size	n = 111229; % = 7
Antidepressant use - SSRIs Sample size	n = 55080; % = 3.5
Antidepressant use - TCAs Sample size	n = 56779; % = 3.6
Antidepressant use - Other Sample size	n = 9976; % = 0.6
Corticosteroid use Sample size	n = 30998; % = 2
Hormone replacement therapy Unopposed Sample size	n = 14988; % = 0.9

Characteristic	Study (N = 1583373)
Care or nursing home resident Sample size	n = 1535; % = 0.1

Outcomes

QFracture (2012 version) outcomes in men

Outcome	Major osteoporotic fracture in men vs No major osteoporotic fracture in men, N2 = 771802, N1 = 7008	Hip fracture in men vs No hip fracture in men, N2 = 776289, N1 = 2521
R^2 (R-squared) (%) Custom value	38.20 (95%CI 36.89-39.57)	70.37 (95%CI 69.25-71.49)
O:E Ratio Custom value	0.866 (95%CI 0.841-0.891)	0.785 (95%CI 0.732-0.839)
AUC Custom value	0.711 (95%CI 0.703-0.719)	0.875 (95%CI 0.868-0.883)
D-statistic Custom value	1.61 (95%CI 1.56-1.66)	3.15 (95%CI 3.06-3.24)
Sensitivity (%) 90% fracture risk threshold. Specificity was not reported nor estimable from reported data.	37.0 (95%CI 36.0-38.0)	64.0 (95%CI 62.0-67.0)

Outcome	Major osteoporotic fracture in men vs No major osteoporotic fracture in men, N2 = 771802, N1 = 7008	Hip fracture in men vs No hip fracture in men, N2 = 776289, N1 = 2521
Custom value		

R^2 (R-squared) - Polarity - Higher values are better

AUC - Polarity - Higher values are better

D-statistic - Polarity - Higher values are better

Sensitivity - Polarity - Higher values are better

n1=fractured group, n2=non-fractured group.

QFracture (2012 version) outcomes in women

Outcome	Major osteoporotic fracture in women vs No major osteoporotic fracture in women, N2 = 771802, N1 = 21677	Hip fracture in women vs No hip fracture in women, N2 = 797474, N1 = 7089
R^2 (R-squared) (%) Custom value	51.9 (95%CI 51.2-52.6)	71.73 (95%CI 71.0-72.30)
O:E Ratio Custom value	0.897 (95%CI 0.876-0.917)	0.799 (95%CI 0.749-0.850)
AUC Custom value	0.790 (95%CI 0.787-0.793)	0.893 (95%CI 0.890-0.896)
D-statistic	2.13 (95%CI 2.10-2.15)	3.26 (95%CI 3.21-3.31)

Outcome	Major osteoporotic fracture in women vs No major osteoporotic fracture in women, N2 = 771802, N1 = 21677	Hip fracture in women vs No hip fracture in women, N2 = 797474, N1 = 7089
Custom value		
Sensitivity (%) 90% fracture risk threshold. Specificity was not reported nor estimable from reported data. Custom value	35.0 (95%CI 34.0-36.0)	60.0 (95%CI 58.0-61.0)

R^2 (R-squared) - Polarity - Higher values are better

AUC - Polarity - Higher values are better

D-statistic - Polarity - Higher values are better

Sensitivity - Polarity - Higher values are better

n1=fractured group, n2=non-fractured group.

Critical appraisal – PROBAST tool for risk prediction models

Overall Risk of bias and Applicability	Risk of bias	High *(QFracture 2012: High RoB for predictors (diagnosis of disease/condition likely to have been variable across participants), outcome (composite fracture outcome using GP records or linked cause of death data) and analysis (Kaplan-Meier analysis used, does not account for competing risks) domains.)*
Overall Risk of bias and Applicability	Directness	Partially applicable *(includes participants not at suspected risk or not at risk of fragility fracture; women < 65 and men < 75 years)*

D.1.6 Hippisley-Cox, 2014

Bibliographic Reference	Hippisley-Cox, Julia; Coupland, Carol; Brindle, Peter; The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study.; BMJ open; 2014; vol. 4 (no. 8); e005809

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study included in review	Original development and internal validation study of QFracture 2012: Hippisley-Cox, J., & Coupland, C. (2012). Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study. *Bmj*, *344*.
Study type	External validation study
Study location	UK
Study setting	Primary care
Study dates	01/01/1998 to 31/07/2012
Sources of funding	National Institute for Health Research School for Primary Care Research (Project reference number 094)
Study sample	External validation cohort from CPRD database, N=2,852,381, from 357 GP practices in England with linked ONS mortality and HES data

Inclusion criteria	- Patient is registered at GP practice on the Clinical Research Data Link Database (CPRD) - Patient aged between 30-99 years-old - GP practice has linked ONS mortality and hospital admissions data
Exclusion criteria	- Patient does not have Townsend score - Patient is temporary resident of GP practice on CPRD database
Risk prediction tool	QFracture 2012 version
Predictors	QFracture 2012 See list of predictors for Hippisley-Cox 2012.
Risk prediction model validation	External validation of 7 QResearch-based prediction models using patients registered on CPRD with data extracted for all predictor variables using same definitions as used in original QFracture 2012 study. Open cohort of patients aged 25-99 years-old at entry to cohort and followed up until 31/07/2015. CPRD cohort restricted to 357 GPs, in England, which had linked ONS mortality and hospital admissions data. For each patient, it was determined when they entered cohort (latest of: 25th birthday, GP registration date plus 1 year, date when GP computer system was installed plus 1 year, beginning of study period 01/01/1998). Patients censored at earliest data of relevant outcome, deregistration with GP practice, last upload of computerised data or study end date (31/12/2015). Eligibility of patients for QFracture-2012 validation then determined. Missing values for BMI, systolic blood pressure, smoking status, alcohol use, and total and HDL cholesterol replaced using multiple imputation by chained equations (5 imputed datasets combined using Rubin's rules to combined effect estimates and standard errors). - Calibration assessed by: calibration plots, mean predicted and observed 10-year risk using Kaplan-Meier estimate compared by deciles of risk; observed risk in top decile of predicted risk - Overall fit assessed by: R^2 - Discrimination assessed by: AUC, and D statistic; sensitivity and specificity in top decile of predicted risk
Outcome	- 10-year risk of osteoporotic fracture - 10-year risk of hip fracture

	Hip fracture identified by presence of READ code on GP record or ICD-10 codes recorded on linked mortality record; osteoporotic fracture identified by READ code on GP record.
Duration of follow-up	Study cohort began 01/01/1998 and was followed up until 31/07/2012

Characteristics

Study-level characteristics

Characteristic	**Study (N = 3271512)**
% Female Data is for total eligible participants in CPRD cohort. Sample size	n = 1682709; % = 51.4
Previous fracture Sample size	n = 70017; % = 2.1
Fall history Sample size	n = 90783; % = 2.8
Cardiovascular disease Sample size	n = 184597; % = 5.6
Cancer Sample size	n = 70774; % = 2.2

Characteristic	Study (N = 3271512)
Chronic liver disease Includes pancreatitis Sample size	n = 9572; % = 0.3
Chronic renal disease Sample size	n = 8050; % = 0.2
Chronic obstructive airways disease or asthma Sample size	n = 312477; % = 9.6
Dementia Sample size	n = 23320; % = 0.7
Endocrine disorders Sample size	n = 177179; % = 0.5
Gastrointestinal malabsorption Sample size	n = 16718; % = 0.5
Parkinson's disease Sample size	n = 9222; % = 0.3
Type 1 diabetes Sample size	n = 11162; % = 0.3

Characteristic	Study (N = 3271512)
Type 2 diabetes Sample size	n = 94905; % = 2.9
Antidepressant use Sample size	n = 337350; % = 10.3
Corticosteroid use Sample size	n = 116949; % = 3.6
Hormone replacement therapy Oestrogen-only HRT Sample size	n = 119413; % = 3.7

Arm-level characteristics

Characteristic	Men (N = 1588803)	Women (N = 1682709)
Mean age (SD) Sample size	n = NA; % = NA	n = NA; % = NA
Mean age (SD) - 25-34 Sample size	n = 427975; % = 26.9	n = 467192; % = 27.8
Mean age (SD) - 35-44	n = 396680; % = 25	n = 364150; % = 21.6

Characteristic	Men (N = 1588803)	Women (N = 1682709)
Sample size		
Mean age (SD) - 45-54 Sample size	n = 294274; % = 18.5	n = 277663; % = 16.5
Mean age (SD) - 55-64 Sample size	n = 212817; % = 13.4	n = 211636; % = 12.6
Mean age (SD) - 65-74 Sample size	n = 148180; % = 9.3	n = 164172; % = 9.8
Mean age (SD) - 75+ Sample size	n = 108877; % = 6.9	n = 197896; % = 11.8
Ethnicity Sample size	n = 587879; % = 37	n = 658835; % = 39.2
Ethnicity - White or not recorded Sample size	n = 1515113; % = 95.4	n = 1602212; % = 95.2
Ethnicity - Indian Sample size	n = 16442; % = 1	n = 16025; % = 1
Ethnicity - Pakistani	n = 6606; % = 0.4	n = 6146; % = 0.4

Characteristic	Men (N = 1588803)	Women (N = 1682709)
Sample size		
Ethnicity - Bangladeshi Sample size	n = 2419; % = 0.2	n = 1688; % = 0.1
Ethnicity - Other Asian Sample size	n = 10795; % = 0.7	n = 11873; % = 0.7
Ethnicity - Caribbean Sample size	n = 4989; % = 0.3	n = 6425; % = 0.4
Ethnicity - Black African Sample size	n = 12883; % = 0.8	n = 14771; % = 0.9
Ethnicity - Chinese Sample size	n = 2914; % = 0.2	n = 4176; % = 0.2
Ethnicity - Another ethnic group Sample size	n = 16642; % = 1	n = 19393; % = 1.2
BMI (kg/m^2) Data for men is for N=1,268,235 (79.8% of eligible cohort); for women, N=1,481,918 (88.1% of eligible cohort). **Mean (SD)**	29.6 (6.8)	28.2 (7)

Characteristic	Men (N = 1588803)	Women (N = 1682709)
Alcohol intake Sample size	n = 1238110; % = 77.9	n = 1379002; % = 82
Alcohol intake - Non-drinker Sample size	n = 163633; % = 10.3	n = 318880; % = 19
Alcohol intake - <1 unit/day Sample size	n = 460091; % = 29	n = 726851; % = 43.2
Alcohol intake - 1-2 units/day Sample size	n = 411261; % = 25.9	n = 290547; % = 17.3
Alcohol intake - 3-6 units/day Sample size	n = 166328; % = 10.5	n = 36763; % = 2.2
Alcohol intake - 7-9 units/day Sample size	n = 19612; % = 1.2	n = 2842; % = 0.2
Alcohol intake - >9 units/day Sample size	n = 17185; % = 1.1	n = 3108; % = 0.2
Smoking status Sample size	n = 1442088; % = 90.5	n = 1595538; % = 94.8

Characteristic	Men (N = 1588803)	Women (N = 1682709)
Smoking status - Non-smoker Sample size	n = 613833; % = 38.6	n = 834721; % = 49.6
Smoking status - Ex-smoker Sample size	n = 252873; % = 15.9	n = 222615; % = 13.2
Smoking status - Light smoker 1-9 cigarettes/day Sample size	n = 104466; % = 6.6	n = 109864; % = 6.5
Smoking status - Moderate smoker 10-19 cigarettes/day Sample size	n = 183000; % = 11.5	n = 179391; % = 10.7
Smoking status - Heavy smoker 20+ cigarettes/day Sample size	n = 142438; % = 9	n = 87474; % = 5.2
Parental history of osteoporosis Family history Sample size	n = 880; % = 0.1	n = 10062; % = 0.6
Family history of - coronary heart disease Sample size	n = 68805; % = 4.3	n = 80985; % = 4.8

Characteristic	Men (N = 1588803)	Women (N = 1682709)
Family history of - Diabetes Sample size	n = 96810; % = 6.1	n = 132390; % = 7.9
Family history of - kidney disease Sample size	n = 1253; % = 0.1	n = 1586; % = 0.1

Outcomes

QFracture (2012 version) outcomes in adults

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 2806684, N1 = 45697	Hip fracture vs No hip fracture, N2 = 2829141, N1 = 23240
Sensitivity (%) 90% fracture risk threshold Custom value	50.0 (95%CI 49.0-50.0)	67.0 (95%CI 66.0-67.0)
Specificity (%) 90% fracture risk threshold Custom value	90.0 (95%CI 90.0-91.0)	90.0 (95%CI 90.0-91.0)

Sensitivity - Polarity - Higher values are better

Specificity - Polarity - Higher values are better

n1=fractured group, n2=non-fractured group

QFracture (2012 version) outcomes in men

Outcome	Major osteoporotic fracture in men vs No major osteoporotic fracture in men, N2 = 1577634, N1 = 11169	Hip fracture in men vs No hip fracture in men, N2 = 1583096, N1 = 5707
R^2 (R-squared) (%) Custom value	49.8 (95%CI 48.9-50.7)	69.0 (95%CI 68.5-70.0)
O:E Ratio O:E ratio and 95% CIs calculated using equations in Table 2, Debray 2018 Custom value	0.744 (95%CI 0.722-0.766)	0.765 (95%CI 0.712-0.817)
AUC Custom value	0.768 (95%CI 0.763-0.773)	0.872 (95%CI 0.867-0.877)
D-statistic Custom value	2.038 (95%CI 2.002-2.075)	2.046 (95%CI 1.977-2.116)

R^2 (R-squared) - Polarity - Higher values are better

AUC - Polarity - Higher values are better

D-statistic - Polarity - Higher values are better

n1=fractured group, n2=non-fractured group

QFracture (2012 version) outcomes in women

Outcome	Major osteoporotic fracture in women vs No major osteoporotic fracture in women, N2 = 1648181, N1 = 34528	Hip fracture in women vs No hip fracture in women, N2 = 1665176, N1 = 17533
R^2 (R-squared) (%) Custom value	56.3 (95%CI 55.8-56.7)	70.6 (95%CI 70.2-71.0)
O:E Ratio O:E ratio and 95% CIs calculated using equations in Table 2, Debray 2018 Custom value	0.823 (95%CI 0.807-0.839)	0.859 (95%CI 0.805-0.912)
AUC Custom value	0.817 (95%CI 0.814-0.819)	0.890 (95%CI 0.888-0.892)
D-statistic Custom value	2.322 (95%CI 2.301-2.343)	3.171 (95%CI 3.139-3.203)

R^2 (R-squared) - Polarity - Higher values are better

AUC - Polarity - Higher values are better

D-statistic - Polarity - Higher values are better

n1=fractured group, n2=non-fractured group

Critical appraisal – PROBAST tool for risk prediction models

Overall Risk of bias and Applicability	Risk of bias	High *(QFracture 2012: Unclear ROB for participant (excludes people with missing Townsend score) domain; high ROB for predictors (disease status confirmation likely to be variable across participants) and outcome*

		(composite fracture outcome) and analysis (Kaplan-Meier analysis used, does not account for competing risks) domains.)
Overall Risk of bias and Applicability	Directness	Partially applicable *(includes participants not at suspected risk or not at risk of fragility fracture; women < 65 and men < 75 years)*

D.1.7 Ihama, 2021

Bibliographic Reference	Ihama, F; Pandyan, A; Roffe, C; Assessment of fracture risk tools in care home residents: a multi-centre observational pilot study.; European geriatric medicine; 2021; vol. 12 (no. 1); 79-89

Study details

Secondary publication of another included study- see primary study for details	NA
Study type	External validation study Prospective cohort study
Study location	Boston, Lincolnshire, UK
Study setting	Care homes (13 residential care homes, 4 nursing homes, 1 home for adults with disability)
Study dates	Not reported, ethical approval obtained 01/2015.
Sources of funding	None. Study was for main author's Doctor of Medicine Thesis, Keele University, UK.

Study sample	External validation cohort, N=217, from 18 UK care homes
Inclusion criteria	- Resident of one of the 18 care homes (residential homes [social and personal care], nursing homes [social, personal, and 24-hour nursing care], and adult disability homes [residential homes for adults with learning disabilities]) in Boston, Lincolnshire, UK - Informed written consent obtained from either resident if mentally competent or consultee (person who is empowered with Lasting Power of Attorney) if applicable
Exclusion criteria	- Resident on end-of-life care pathway - Informed written consent not obtainable from either resident or consultee
Population subgroups	NA
Risk prediction	FRAX-UK FRAX-UK without BMD QFracture 2016 version
Predictors	**FRAX without BMD** - Age - Sex - Weight - Height - Prior fragility fracture - Parental history of hip fracture - Current tobacco smoking - Long-term use of oral glucocorticoids - Rheumatoid arthritis - Other causes of secondary osteoporosis

	- Daily alcohol consumption of three or more units daily **<u>QFracture 2016</u>** See list of predictors for Hippisley-Cox 2012
Risk prediction model validation	External validation of FRAX-UK without BMD and QFracture 2016 in care home residents. Follow up of 12 months chosen due to high mortality rate of care home residents. Baseline assessment of fracture risk conducted using structured composite questionnaire that captured all covariates in pdf versions of FRAX-UK without BMD, QFracture-2016, Garvan nomogram, BMI, the Timed Up & Go Test (TUGT) falls risk assessment tool, and the Charlson Comorbidity Index (CCI). Fracture confirmed by X-ray and reported as such by radiologist; fragility fracture defined as one sustained after low trauma, quantified by WHO as forces equivalent to fall from standing height or less; skull fractures, facial fractures, and fractures result of road traffic accident and pathological fractures excluded. - Calibration and overall fit statistics not reported. - Discrimination assessed by: AUC *Note: Data extracted only for FRAX-UK without BMD and QFracture 2016.*
Outcome	- 10-year risk of osteoporotic fracture Osteoporotic fractures identified by radiographs evaluated by radiologist.
Duration of follow-up	FU=12 months

Characteristics

Study-level characteristics

Characteristic	Study (N = 217)
% Female	n = 134; % = 61.8
Sample size	

Characteristic	Study (N = 217)
Mean age (SD) Mean (SD)	81.21 (12.51)
Ethnicity Sample size	n = NA; % = NA
Ethnicity - Caucasian Sample size	n = 217; % = 100
Ethnicity - Other Sample size	n = 0; % = 0
BMI (kg/m^2) Mean (SD)	24.26 (7.21)
Alcohol intake Sample size	n = NA; % = NA
Alcohol intake - ≥3 units/day Sample size	n = 5; % = 2.3
Alcohol intake - ≥3 units/day Sample size	n = 212; % = 97.7
Smoking status	n = NA; % = NA

Characteristic	Study (N = 217)
Sample size	
Smoking status - smoker Sample size	n = 9; % = 4.2
Smoking status - Non-smoker Sample size	n = 208; % = 95.8
Nursing or care home resident Sample size	n = 217; % = 100
Nursing or care home resident - Residential Sample size	n = 177; % = 81.6
Nursing or care home resident - Nursing home Sample size	n = 40; % = 18.4
Charlson comorbidity index for 1 year Mean (SD)	30.65 (20.75)

Outcomes

FRAX without BMD outcomes

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 207, N1 = 10
c-statistic 95% CI calculated from reported total number of patients, number of incident fractures, and number of patients not experiencing incident fractures using equations in Table 2, Debray 2018 Custom value	0.655 (95%CI 0.469-0.803)

c-statistic - Polarity - Higher values are better

n1=fractured group, n2=non-fractured group

QFracture (2016 version) outcomes

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 207, N1 = 10
c-statistic 95% CI calculated from reported total number of patients, number of incident fractures, and number of patients not experiencing incident fractures using equations in Table 2, Debray 2018Custom value	0.736 (95%CI 0.553-0.862)

c-statistic - Polarity - Higher values are better

n1=fractured group, n2=non-fractured group

Critical appraisal – PROBAST tool for risk prediction models

Overall Risk of bias and Applicability	Risk of bias	High *(FRAX-UK/QFracture 2016: Unclear RoB for participant (excludes people where consent could not be obtained) domain. High RoB for outcome (Radiologist verified fractures by x-ray, highly subjective) and*

		analysis (no information about missing data strategy; inappropriate calibration statistics; does not account for censoring/competing risks) domains.)
Overall Risk of bias and Applicability	Directness	Low

D.1.8 Klop, 2016

Bibliographic Reference Klop, Corinne; de Vries, Frank; Bijlsma, Johannes W J; Leufkens, Hubert G M; Welsing, Paco M J; Predicting the 10-year risk of hip and major osteoporotic fracture in rheumatoid arthritis and in the general population: an independent validation and update of UK FRAX without bone mineral density.; Annals of the rheumatic diseases; 2016; vol. 75 (no. 12); 2095-2100

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study included in review	NA
Study type	External validation study
Study location	UK

Study setting	Primary care
Study dates	01/01/1987 to 31/12/2013
Sources of funding	Supported by Netherlands Organization for Health Research and Development (ZonMw, grant #113101007)
Study sample	External validation CPRD cohort (M/F) used for recalibration of FRAX-UK, N=11,582 (3729/7853) people with rheumatoid arthritis (RA) External validation CPRD-HES cohort used for extension of FRAX-UK, N=7,221 (2263/4958) people with RA External validation CPRD matched cohort from general population, N=38,755 (N=24,227 with HES linkage) people from general population
Inclusion criteria	- Patient registered at primary care practices included in Clinical Practice Research Datalink (CPRD) - Patient has diagnosis of rheumatoid arthritis at 01/01/2004 (index date), defined as 1 or more disease-modifying antirheumatic drug prescription after first RA diagnosis code - Patient had data collected for at least 1-year before index date - Aged 40-90 years at index date
Exclusion criteria	- Previous exposure to any anti-osteoporosis drug before index date
Risk prediction tool	FRAX-UK FRAX-UK without BMD
Predictors	FRAX-UK without BMD - Age - Sex - Weight - Height

	- Prior fragility fracture - Parental history of hip fracture - Current tobacco smoking - Long-term use of oral glucocorticoids - Rheumatoid arthritis - Other causes of secondary osteoporosis - Daily alcohol consumption of three or more units daily
Risk prediction model validation	Identified people with RA were matched with up to 4 controls from general population by age, sex, and practice. Missing values for BMI, smoking status and alcohol use replaced using multiple imputation using all predictors and outcome variable, resulting in 5 imputed datasets. Predicted 10-year risks of major osteoporotic fracture (MOF) and hip fracture (HF) calculated without BMD for every participant, repeated for each imputed dataset to provide mean predicted risks. Observed 10-year risk of MOF and HF estimated by cumulative incidence function (accounts for competing risks). Fractures assessed between index date and first occurrence of: death, end of study period (truncated at 10 years from index date: 31/12/2013), or moving out of CPRD. - Calibration assessed by: calibration plot, beta-coefficient - Discrimination assessed by: c-statistics. - Reclassification assessed by: net reclassification improvement (NRI), recalibrated FRAX compared to extended-FRAX Sensitivity analysis assessing influence of anti-osteoporosis drug (AOD) use after index date on average observed risk conducted by increasing observed risk inversely proportional to estimated AODs effect on HF. Two additional analyses conducted: Recalibration (update) of FRAX-UK without BMD for hip fracture in CPRD-HES RA cohort conducted by fitting log-odds transformed FRAX probabilities (i.e. the linear predictor) as single continuous covariate in logistic regression model with outcome of hospitalization for HF within 10 years. Extension of FRAX-UK without BMD for hip fracture using CPRD-HES RA cohort conducted by adding individual FRAX predictors, corticosteroid dose, and duration of RA to linear predictor to determine any additional predictive effect. Interactions between linear predictor and FRAX predictors, glucocorticoid dose and RA disease duration also tested. Derivation of updated FRAX model included all variables and interactions significantly related to HF risk in multivariable model and proceeded by backward elimination. Sensitivity analysis conducted including antiosteoporosis drug treatment after index data and its interaction terms in model.

	Comparison of recalibrated and extended FRAX models compared with c-statistic and Net Reclassification Improvement (NRI). NRI incorporates National Osteoporosis Guideline Group (NOGG) age-specific intervention thresholds which are linked to FRAX output in UK. Bootstrapping of 500 repetitions performed to correct c-statistic for optimism. Shrinkage factor applied to beta-coefficients of final model.
Outcome	*People with rheumatoid arthritis* - 10-year risk of major osteoporotic fracture - 10-year risk of hip fracture *General population* - 10-year risk of major osteoporotic fracture - 10-year risk of hip fracture Both types of fracture identified by READ codes in CPRD/CPRD-HES databases.
Duration of follow-up	CPRD RA cohort Median FU=9.0 years (IQR 4.7-10.0) CPRD-HES RA cohort Median FU=9.0 years (IQR 5.3-10.0)

Characteristics

Study-level characteristics

Characteristic	**Study (N = 11582)**
% Female Sample size	n = 7853; % = 67.8

Characteristic	Study (N = 11582)
Mean age (SD) Mean (SD)	62.9 (11.4)
Alcohol intake Sample size	n = NA; % = NA
Alcohol intake - ≥3 units/day Sample size	n = 580; % = 5
Alcohol intake - Not recorded Sample size	n = 1759; % = 15.2
Smoking status Sample size	n = NA; % = NA
Smoking status - Current smoker Sample size	n = 4147; % = 35.8
Smoking status - Not recorded Sample size	n = 890; % = 7.7
Previous fracture Sample size	n = 1908; % = 16.5

Characteristic	Study (N = 11582)
Corticosteroid use Prescription within 90 days before or >2 prescriptions with mean daily dose of prednisolone or equivalents of ≥5 mg in year before. Sample size	n = 1806; % = 15.6
Rheumatoid arthritis disease duration - <2 years since diagnosis Sample size	n = 1336; % = 11.5
Rheumatoid arthritis disease duration - 2-10 years since diagnosis Sample size	n = 5900; % = 50.9
Rheumatoid arthritis disease duration - >10 years since diagnosis Sample size	n = 4346; % = 37.5
Secondary osteoporosis Sample size	n = 580; % = 5

Outcomes

FRAX-UK without BMD outcomes in adults aged 40-90 from primary care population

Outcome	Major osteoporotic fracture in adults vs No major osteoporotic fracture in adults, N2 = 36830, N1 = 1925	Hip fracture in adults vs No hip fracture in adults, N2 = 38219, N1 = 536
O:E Ratio O:E ratio and 95% CIs calculated using equations in Table 2, Debray 2018 Custom value	NR	0.884 (95%CI 0.773-0.995)
c-statistic 95% CI calculated from reported total number of patients, number of incident fractures, and number of patients not experiencing incident fractures using equations in Table 2, Debray 2018 Custom value	0.69 (95%CI 0.671-0.708)	0.83 (95%CI 0.812-0.847)

c-statistic - Polarity - Higher values are better

FRAX-UK without BMD outcomes in adults with rheumatoid arthritis aged 40-90 from primary care population

Outcome	Major osteoporotic fracture in adults with RA vs No major osteoporotic fracture in adults with RA, N2 = 10774, N1 = 808	Hip fracture in adults with RA vs No hip fracture in adults with RA, N2 = 11285, N1 = 297
O:E Ratio O:E ratio calculated using equations, Table 2, Debray 2018 Custom value	0.632 (95%CI 0.558-0.706)	NR
O:E Ratio - Calibration of model to RA population Custom value	NR	0.748 (95%CI 0.561-0.935)

Outcome	Major osteoporotic fracture in adults with RA vs No major osteoporotic fracture in adults with RA, N2 = 10774, N1 = 808	Hip fracture in adults with RA vs No hip fracture in adults with RA, N2 = 11285, N1 = 297
O:E Ratio - Recalibration of model to RA population Custom value	NR	0.748 (95%CI 0.511-0.985)
O:E Ratio - Extension of model to RA population Custom value	NR	0.943 (95%CI 0.649-1.238)
c-statistic 95% CI calculated from reported total number of patients, number of incident fractures, and number of patients not experiencing incident fractures using equations in Table 2, Debray 2018 Custom value	0.69 (95%CI 0.671-0.708)	0.78 (95%CI 0.752-0.805)

Reclassification

Outcome	Extended FRAX-UK vs Recalibrated FRAX-UK, N2 = NA, N1 = NA
Net reclassification index Extended FRAX-UK (includes additional predictors: duration of RA, use of high-dose glucocorticoid, and secondary osteoporosis) compared to recalibrated FRAX-UK using NOGG age-specific thresholds Custom value	0.01 (95% CI -0.04-0.05)

Critical appraisal – PROBAST tool for risk prediction models

Overall Risk of bias and Applicability	Risk of bias	High *(FRAX-UK: High RoB for outcome (Composite outcome using primary care and HES database) domain. Outcome for recalibrated model in RA population changed to hospitalisation for hip fracture within 10 years) and analysis (Calibration plots reported but no other calibration measures) domains)*
Overall Risk of bias and Applicability	Directness	Directly applicable (RA population) Partially applicable (matched cohort) *(includes participants not at suspected risk or not at risk of fragility fracture; women < 65 and men < 75 years)*

D.1.9 Livingstone, Guthrie et al, 2023

Bibliographic Reference Livingstone, Shona J; Guthrie, Bruce; McMinn, Megan; Eke, Chima; Donnan, Peter T; Morales, Daniel R; Derivation and validation of the CFracture competing risk fracture prediction tool compared with QFracture in older people and people with comorbidity: a population cohort study.; The lancet. Healthy longevity; 2023; vol. 4 (no. 1); e43-e53

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study included in review	Livingstone 2022

Study type	Internal validation study
Study location	UK
Study setting	Primary care
Study dates	01/012004 to 31/03/2016
Sources of funding	Study funded by NIHR Health and Social Care Delivery Research Programme (ref: 15/12/22)
Study sample	Derivation (development) cohort, N=3,621,526 (1,789,920/1,831,606) (2/3 of Livingstone 2022 cohort) Internal validation cohort for CFracture, N=1,810,713 (894,910/915,803) (1/3 of Livingstone 2022 cohort) *Note: This study uses the same population cohort as Livingstone 2022 to develop and internally validate CFracture. Results for QFracture 2012 reported in this article were not extracted because they are for a subset of the Livingstone 2022 external validation cohort, which is already included in this review.*
Inclusion criteria	- Patient permanently registered with GP practice contributing up-to-standard data for at least 1 year - Linkage to Hospital Episodes Statistics and ONS mortality data - Aged 30-99 years - Observable records on or after 01/01/2004
Exclusion criteria	NR
Population subgroups	- Men - Women
Risk prediction tool	CFracture
Predictors	**CFracture** - Age

	- Sex - Ethnicity - Smoking status - Alcohol status - Type 1 and type 2 diabetes - BMI - Parental history of osteoporosis/hip fracture - Resident in a nursing or care home - Previous fragility fracture - History of falls - Asthma - Cancer - Cardiovascular disease - Chronic kidney disease - Chronic liver disease - Chronic obstructive pulmonary disease - Dementia - Epilepsy - Gastrointestinal malabsorption (e.g. Crohn's disease, ulcerative colitis, celiac disease, steatorrhoea, blind loop syndrome) - Parkinson's disease - Rheumatoid arthritis - Systemic lupus erythematosus - Other endocrine conditions (thyrotoxicosis, primary or secondary hyperparathyroidism, Cushing's syndrome) - ≥2 prescriptions for systemic corticosteroids in the six months prior to cohort entry - ≥2 prescriptions for antidepressants six months prior to cohort entry - ≥2 prescriptions for oestrogen only hormone replacement therapy (in women) six months prior to cohort entry
Risk prediction model validation	CFracture

	Patients randomly allocated 2:1 ratio to derivation (development) and internal validation dataset using same population cohort used to externally validate QFracture 2012 in Livingstone 2022, with split balanced in terms of age and outcome status. Derivation dataset used to derive CFracture, a Fine-Grey model to predict 10-year MOF or HF accounting for competing risk of non-fracture death with separate models for men and women. Unlike QFracture 2012, data on BMI, alcohol use and smoking status restricted to values recorded before cohort entry. Charlson Comorbidity Index (CCI) additionally calculated on basis of READ codes with CCI category (0, 1, 2, or ≥3) included in competing risk model as predictor of competing mortality risk. Observed risk estimated using Aalen-Johansen estimator, which accounts for competing mortality risk. Individuals with missing data for ethnicity presumed to be white. Missing BMI, smoking status, and alcohol status were imputed using multiple imputation with chained equations to generate 5 imputed datasets combined using Rubin's rules. Missing data for morbidities and prescription medicine assumed to be absent if not recorded. - Calibration assessed by calibration plots; O:E ratio - Discrimination assessed by: c-statistic *Note: Results for QFracture 2012 reported in this article were not extracted because they are for a subset of the Livingstone 2022 external validation cohort, which is already included in this review.*
Outcome	- 10-year risk of major osteoporotic (hip, vertebral, wrist, proximal humerus) fracture - 10-year risk of hip fracture Both outcomes identified using READ codes from CPRD GOLD database, HES discharge diagnoses, or ONS death registration (ICD-10) codes.
Duration of follow-up	Reported as 10,419,774 person-years of follow up for major osteoporotic fracture and 11,446,139 person-years of follow up for hip fracture in derivation cohort. Follow up not reported for validation cohort.

Characteristics

Study-level characteristics

Characteristic	Study (N = 1810713)
% Female Sample size	n = 915803; % = 50.6

Arm-level characteristics

Characteristic	Men (N = 894910)	Women (N = 915803)
Mean age (SD) Median (IQR)	45 (35 to 59)	47 (35 to 63)
Ethnicity Sample size	n = NA; % = NA	n = NA; % = NA
Ethnicity - White or not recorded Sample size	n = 852170; % = 95.2	n = 871830; % = 95.2
Ethnicity - Indian Sample size	n = 9062; % = 1	n = 8355; % = 0.9
Ethnicity - Pakistani Sample size	n = 4072; % = 0.5	n = 3719; % = 0.4
Ethnicity - Bangladeshi	n = 1663; % = 0.2	n = 1087; % = 0.1

Characteristic	Men (N = 894910)	Women (N = 915803)
Sample size		
Ethnicity - Black African Sample size	n = 7007; % = 0.8	n = 7595; % = 0.8
Ethnicity - Black Caribbean Sample size	n = 1332; % = 0.1	n = 1633; % = 0.2
Ethnicity - Chinese Sample size	n = 1824; % = 0.2	n = 2398; % = 0.3
Ethnicity - Another Asian ethnic group Sample size	n = 5846; % = 0.7	n = 6302; % = 0.7
Ethnicity - Another ethnic group Sample size	n = 11934; % = 1.3	n = 12884; % = 1.4
BMI (kg/m^2) Median (IQR)	26.5 (23.9 to 29.5)	25.4 (22.4 to 29.6)
Alcohol intake Percentages are for patients who reported alcohol use (denominator does not include patients who did not report alcohol use) Sample size	n = NA; % = NA	n = NA; % = NA

Characteristic	Men (N = 894910)	Women (N = 915803)
Alcohol intake - None Sample size	n = 106229; % = 17.5	n = 190034; % = 27.8
Alcohol intake - <1 unit/day Sample size	n = 182618; % = 30.2	n = 284482; % = 41.7
Alcohol intake - 1-2 units/day Sample size	n = 222747; % = 36.8	n = 187779; % = 27.5
Alcohol intake - 3-6 units/day Sample size	n = 74712; % = 12.3	n = 17548; % = 2.6
Alcohol intake - 7-9 units/day Sample size	n = 12716; % = 2.1	n = 1973; % = 0.3
Alcohol intake - >9 units/day Sample size	n = 6629; % = 1.1	n = 947; % = 0.1
Alcohol intake - Not recorded Sample size	n = 289259; % = 32.3	n = 233041; % = 25.4
Smoking status Percentages are for patients who reported smoking status (denominator does not include patients who did not report smoking status)	n = NA; % = NA	n = NA; % = NA

Characteristic	Men (N = 894910)	Women (N = 915803)
Sample size		
Smoking status - Non-smoker Sample size	n = 355153; % = 53.8	n = 494879; % = 64.4
Smoking status - Ex-smoker Sample size	n = 146333; % = 22.2	n = 130128; % = 16.9
Smoking status - Light smoker <10 cigarettes/day Sample size	n = 41684; % = 6.3	n = 45377; % = 5.9
Smoking status - Moderate smoker 10-19 cigarettes/day Sample size	n = 63797; % = 9.7	n = 62705; % = 8.2
Smoking status - Heavy smoker ≥20 cigarettes/day Sample size	n = 52860; % = 8	n = 35629; % = 4.6
Smoking status - Not recorded Sample size	n = 235083; % = 26.3	n = 147085; % = 16.1
Previous fracture Previous major osteoporotic fracture	n = 37801; % = 4.2	n = 50095; % = 5.5

Characteristic	Men (N = 894910)	Women (N = 915803)
Sample size		
Parental history of osteoporosis Parental history of osteoporosis or hip fracture Sample size	n = 347; % = 0.1	n = 3621; % = 0.4
Fall history Sample size	n = 24996; % = 2.8	n = 51214; % = 5.6
Cardiovascular disease Sample size	n = 64947; % = 7.3	n = 52466; % = 5.7
Cancer Sample size	n = 22276; % = 2.5	n = 31486; % = 3.4
Chronic liver disease Sample size	n = 2292; % = 0.3	n = 2021; % = 0.2
Chronic renal disease Sample size	n = 8206; % = 0.9	n = 11179; % = 1.2
Chronic obstructive airways disease or asthma Sample size	n = 101270; % = 11.3	n = 118469; % = 12.9
Dementia	n = 4987; % = 0.6	n = 11635; % = 1.3

Characteristic	Men (N = 894910)	Women (N = 915803)
Sample size		
Endocrine disorders Sample size	n = 1937; % = 0.2	n = 8297; % = 0.9
Epilepsy Epilepsy or prescribed anticonvulsants Sample size	n = 19660; % = 2.2	n = 22216; % = 2.4
Gastrointestinal malabsorption Sample size	n = 8981; % = 1	n = 11687; % = 1.3
Parkinson's disease Sample size	n = 2699; % = 0.3	n = 2524; % = 0.3
Rheumatoid arthritis or systemic lupus erythematosus Sample size	n = 3986; % = 0.4	n = 10913; % = 1.2
Type 1 diabetes Sample size	n = 3891; % = 0.4	n = 2938; % = 0.3
Type 2 diabetes Sample size	n = 33424; % = 3.7	n = 27202; % = 3

Characteristic	Men (N = 894910)	Women (N = 915803)
Antidepressant use Sample size	n = 47328; % = 5.3	n = 99744; % = 10.9
Corticosteroid use Sample size	n = 7604; % = 0.8	n = 12378; % = 1.4
Nursing or care home resident Sample size	n = 2461; % = 0.3	n = 5654; % = 0.6

Outcomes

CFracture outcomes in men

Outcome	Major osteoporotic fracture in men vs No major osteoporotic fracture in men, N2 = NR, N1 = NR	Hip fracture in men vs No hip fracture in men, N2 = NR, N1 = NR
O:E Ratio Custom value	1.06 (95%CI 0.98-1.15)	1.18 (95%CI 1.05-1.32)
c-statistic Custom value	0.738 (95%CI 0.732-0.743)	0.886 (95%CI 0.877-0.895)

c-statistic - Polarity - Higher values are better

CFracture outcomes in women

Outcome	Major osteoporotic fracture in women vs No major osteoporotic fracture in women, N2 = NR, N1 = NR	Hip fracture in women vs No hip fracture in women, N2 = NR, N1 = NR
O:E Ratio Custom value	1.16 (95%CI 1.11-1.21)	1.07 (95%CI 0.98-1.15)
c-statistic Custom value	0.813 (95%CI 0.810-0.816)	0.914 (95%CI 0.908-0.919)

c-statistic - Polarity - Higher values are better

Critical appraisal – PROBAST tool for risk prediction models

Overall Risk of bias and Applicability	Risk of bias	High *(CFracture: High RoB on predictor (disease status diagnosis likely to be variable) and outcome (composite fracture outcome) domain)*
Overall Risk of bias and Applicability	Directness	Partially applicable *(includes participants not at suspected risk or not at risk of fragility fracture; women < 65 and men < 75 years)*

D.1.10 Livingstone, Morales et al, 2022

Bibliographic Reference Livingstone, Shona J; Morales, Daniel R; McMinn, Megan; Eke, Chima; Donnan, Peter; Guthrie, Bruce; Effect of competing mortality risks on predictive performance of the QFracture risk prediction tool for major osteoporotic fracture and hip fracture: external validation cohort study in a UK primary care population.; BMJ medicine; 2022; vol. 1 (no. 1); e000316

Study details

Secondary publication of	NA

another included study- see primary study for details	
Other publications associated with this study included in review	Original development and internal validation study for QFracture-2012: Hippisley-Cox, J., & Coupland, C. (2012). Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study. *Bmj*, *344*
Study type	External validation study
Study location	UK
Study setting	Primary care
Study dates	01/01/2004 to 31/03/2016
Sources of funding	Funded by National Institute for Health Research (NIHR) Health Services and Delivery Research Programme (ref: 15/12/22).
Study sample	External validation cohort (M/F), N=5,432,139 (2,684,730/2,747,409) from Clinical Practice Research Datalink (CPRD)-Gold database
Inclusion criteria	- Patient permanently registered with GP practice included in CPRD GOLD primary care database and that contributed at least 1 year up-to-standard data (i.e. GPs pass CPRD-Gold quality checks) - Patient data linked to Hospital Episodes Statistics discharge data and ONS mortality data - Aged 30 to <100 years-old
Population subgroups	- Men - Women

Risk prediction tool	QFracture 2012 version *Note: original version of article states that this is validation of 2016 version, but this has subsequently been corrected (see https://bmjmedicine.bmj.com/content/2/1/e000316corr1)*
Predictors	For list of QFracture 2012 predictors, see entry for Hippisley-Cox 2012. Derived model used in this study, unlike QFracture-2012, only allowed predictor values recorded before study entry in risk prediction.
Risk prediction model validation	Outcomes identified using GP health record READ codes, HES discharge diagnoses (ICD-10 codes as reason for hospital admission), and ONS death registration (ICD-10) codes. Codes used by QFracture not published so authors derived their own for this study. 10-year predicted risk of major osteoporotic fracture and hip fracture calculated using published QFracture-2012 model for all patients in cohort. Charlson Comorbidity Index calculated for each patient at baseline, which was used to classify discrimination and calibration analysis by comorbidity level (it was not used in fracture prediction). Missing data on BMI, smoking status and alcohol use replaced using multiple (multivariate) imputation by chained equations. This generated 5 imputed datasets that were combined using Rubin's rules; ethnicity assumed to be white if missing; morbidities and prescribing use were assumed to be absent if no relevant data recorded. Calibration of QFracture 2012 (without competing mortality risks) assessed by: plotting observed proportions of fractures and predicted probabilities of fracture by deciles of risk. Observed risk for censored data estimated using standard Kaplan-Meier estimator (does not account for competing risks) Calibration of QFracture 2012 extended model (with competing mortality risks) assessed by: plotting observed proportions of fractures and predicted probabilities of fracture by deciles of risk using Aalen Johansen estimator (accounts for competing events, in this case death from causes other than fractures) Overall fit assessed by: R^2 Discrimination assessed by: Harrell's c-statistic (shortened to include only pairs where earliest survival time is no later than 10 years after entry), Royston & Sauerbrei's D statistic.
Outcome	- 10-year risk of major osteoporotic (hip, vertebral, wrist, or proximal humeral fracture) fracture - 10-year risk of hip fracture Both outcomes identified using GP health record READ codes, HES discharge diagnoses, and ONS death registration.

Duration of follow-up	For MOF in men, median FU=5.6 years (IQR 2.2 to 10.4) For MOF in women, median FU=5.7 years (IQR 2.2 to 10.5) For HF in men, median FU=5.7 years (IQR 2.2 to 10.4) For HF in women, median FU=5.9 years (IQR 2.2 to 10.6)

Characteristics

Study-level characteristics

Characteristic	**Study (N = 5432139)**
% Female Sample size	n = 2747409; % = 50.6

Arm-level characteristics

Characteristic	**Men (N = 2684730)**	**Women (N = 2747409)**
Mean age (SD) (years) Mean (SD)	48.5 (15.6)	50.7 (17.4)
Ethnicity Sample size	n = NA; % = NA	n = NA; % = NA
Ethnicity - White or not recorded	n = 2556923; % = 95.2	n = 2614423; % = 95.2

Characteristic	Men (N = 2684730)	Women (N = 2747409)
Sample size		
Ethnicity - Indian	n = 27087; % = 1	n = 25420; % = 0.9
Sample size		
Ethnicity - Pakistani	n = 12316; % = 0.5	n = 11121; % = 0.4
Sample size		
Ethnicity - Bangladeshi	n = 4972; % = 0.2	n = 3473; % = 0.1
Sample size		
Ethnicity - Other Asian	n = 17758; % = 0.7	n = 18896; % = 0.7
Sample size		
Ethnicity - Black Caribbean	n = 4030; % = 0.2	n = 4780; % = 0.2
Sample size		
Ethnicity - Black African	n = 20776; % = 0.8	n = 22736; % = 0.8
Sample size		
Ethnicity - Chinese	n = 5517; % = 0.2	n = 7358; % = 0.3
Sample size		
Ethnicity - Another ethnic group	n = 35351; % = 1.3	n = 392021; % = 1.4
Sample size		

Characteristic	Men (N = 2684730)	Women (N = 2747409)
BMI (kg/m^2) Sample size	n = 27.1; % = 4.8	n = 26.6; % = 6
Alcohol intake Sample size	n = NA; % = NA	n = NA; % = NA
Alcohol intake - None Sample size	n = 317208; % = 11.8	n = 570900; % = 20.8
Alcohol intake - <1 unit/day Sample size	n = 548761; % = 20.4	n = 854476; % = 31.1
Alcohol intake - 1-2 units/day Sample size	n = 669776; % = 24.9	n = 561603; % = 20.4
Alcohol intake - 3-6 units/day Sample size	n = 224507; % = 8.4	n = 52785; % = 1.9
Alcohol intake - 7-9 units/day Sample size	n = 38273; % = 1.4	n = 5750; % = 0.2
Alcohol intake - >9 units/day Sample size	n = 9583; % = 0.7	n = 2993; % = 0.1
Alcohol intake - Not recorded	n = 866622; % = 32.3	n = 698902; % = 25.4

Characteristic	Men (N = 2684730)	Women (N = 2747409)
Sample size		
Smoking status Sample size	n = NA; % = NA	n = NA; % = NA
Smoking status - Non-smoker Sample size	n = 807294; % = 30.1	n = 1146025; % = 41.7
Smoking status - Ex-smoker Sample size	n = 439503; % = 16.4	n = 390520; % = 14.2
Smoking status - Light smoker <10 cigarettes/day Sample size	n = 125229; % = 4.7	n = 135272; % = 4.9
Smoking status - Moderate smoker 10-19 cigarettes/day Sample size	n = 190990; % = 7.1	n = 188078; % = 6.8
Smoking status - Heavy smoker >20 cigarettes/day Sample size	n = 158134; % = 5.9	n = 107288; % = 3.9
Smoking status - Current smoking amount not recorded Sample size	n = 78372; % = 2.9	n = 43957; % = 1.6

Characteristic	Men (N = 2684730)	Women (N = 2747409)
Smoking status - Not recorded Sample size	n = 963580; % = 33	n = 780226; % = 26.8
Previous fracture Previous major osteoporotic fracture Sample size	n = 113520; % = 4.2	n = 152417; % = 5.5
Parental history of osteoporosis Parental history of osteoporosis or hip fracture Sample size	n = 1007; % = 0.0004	n = 10561; % = 0.4
Fall history Sample size	n = 74368; % = 2.8	n = 153841; % = 5.6
Cardiovascular disease Sample size	n = 195378; % = 7.3	n = 156577; % = 5.7
Cancer Sample size	n = 67380; % = 2.5	n = 94090; % = 3.4
Chronic liver disease Sample size	n = 6753; % = 0.3	n = 6093; % = 0.2
Chronic renal disease Sample size	n = 24395; % = 0.9	n = 33274; % = 1.2

Characteristic	Men (N = 2684730)	Women (N = 2747409)
Chronic obstructive airways disease or asthma Sample size	n = 303541; % = 11.3	n = 355014; % = 12.9
Dementia Sample size	n = 15036; % = 0.6	n = 34892; % = 1.3
Endocrine disorders Sample size	n = 5866; % = 0.2	n = 25089; % = 0.9
Epilepsy Epilepsy or prescribed anticonvulsants Sample size	n = 59214; % = 2.2	n = 66145; % = 2.4
Gastrointestinal malabsorption Sample size	n = 27122; % = 1	n = 34884; % = 1.3
Parkinson's disease Sample size	n = 8348; % = 0.3	n = 7585; % = 0.3
Rheumatoid arthritis or systemic lupus erythematosus Sample size	n = 32950; % = 1.2	n = 11970; % = 0.4
Type 1 diabetes Sample size	n = 12008; % = 0.4	n = 8747; % = 0.3

Characteristic	Men (N = 2684730)	Women (N = 2747409)
Type 2 diabetes Sample size	n = 100009; % = 3.7	n = 81715; % = 3
Antidepressant use Sample size	n = 59214; % = 2.2	n = 66145; % = 2.4
Corticosteroid use Sample size	n = 22632; % = 0.8	n = 37169; % = 1.4
Hormone replacement therapy Oestrogen only HRT Sample size	n = 127; % = 0	n = 33679; % = 1.2
Nursing or care home resident Sample size	n = 7455; % = 0.3	n = 16819; % = 0.6

Outcomes

QFracture (2012 version) outcomes for men

Outcome	Major osteoporotic fracture in men vs No major osteoporotic fracture in men, N2 = 2650409, N1 = 34321	Hip fracture in men vs No hip fracture in men, N2 = 2671351, N1 = 13379
R^2 (R-squared) (%) Custom value	42.4 (95%CI 41.9-43.0)	70.9 (95%CI 70.4-71.3)

Outcome	Major osteoporotic fracture in men vs No major osteoporotic fracture in men, N2 = 2650409, N1 = 34321	Hip fracture in men vs No hip fracture in men, N2 = 2671351, N1 = 13379
O:E Ratio Custom value	NA	NA
O:E Ratio - Kaplan-Meier estimator Original QFracture 2012 model, does not account for competing risks Custom value	1.817 (95%CI 1.806-1.827)	1.757 (95%CI 1.720-1.793)
O:E Ratio - Aalen Johansen estimator Alternative QFracture 2012 model, accounts for competing risks (deaths from non-fracture causes) Custom value	1.483 (95%CI 1.473-1.494)	1.319 (95%CI 1.288-1.349)
c-statistic Custom value	0.738 (95%CI 0.735-0.741)	0.888 (95%CI 0.882-0.893)
D-statistic Custom value	1.76 (95%CI 1.74-1.78)	3.19 (95%CI 3.16-3.23)

R^2 (R-squared) - Polarity - Higher values are better

c-statistic - Polarity - Higher values are better

D-statistic - Polarity - Higher values are better

QFracture (2012 version) outcomes for women

Outcome	Major osteoporotic fracture in women vs No major osteoporotic fracture in women, N2 = 2651811, N1 = 95598	Hip fracture in women vs No hip fracture in women, N2 = 2711009, N1 = 36400
R^2 (R-squared) (%) Custom value	54.8 (95%CI 54.5-55.1)	71.7 (95%CI 71.4-71.9)
O:E Ratio Custom value	NA	NA
O:E Ratio - Kaplan-Meier estimator Original QFracture 2012 model, does not account for competing risks. Insufficient data reported to calculate 95%CIs for hip fracture Custom value	1.508 (95%CI 1.481-1.536)	1.306
O:E Ratio - Aalen Johansen estimator Alternative QFracture 2012 model, accounts for competing risks (deaths from non-fracture causes). Insufficient data reported to calculate 95%CIs for hip fracture Custom value	1.212 (95%CI 1.185-1.239)	0.930
c-statistic Custom value	0.813 (95%CI 0.811-0.815)	0.918 (95%CI 0.915-0.921)
D-statistic Custom value	2.25 (95%CI 2.24-2.27)	3.26 (95%CI 3.24-3.28)

R^2 (R-squared) - Polarity - Higher values are better

c-statistic - Polarity - Higher values are better

D-statistic - Polarity - Higher values are better

Critical appraisal – PROBAST tool for risk prediction models

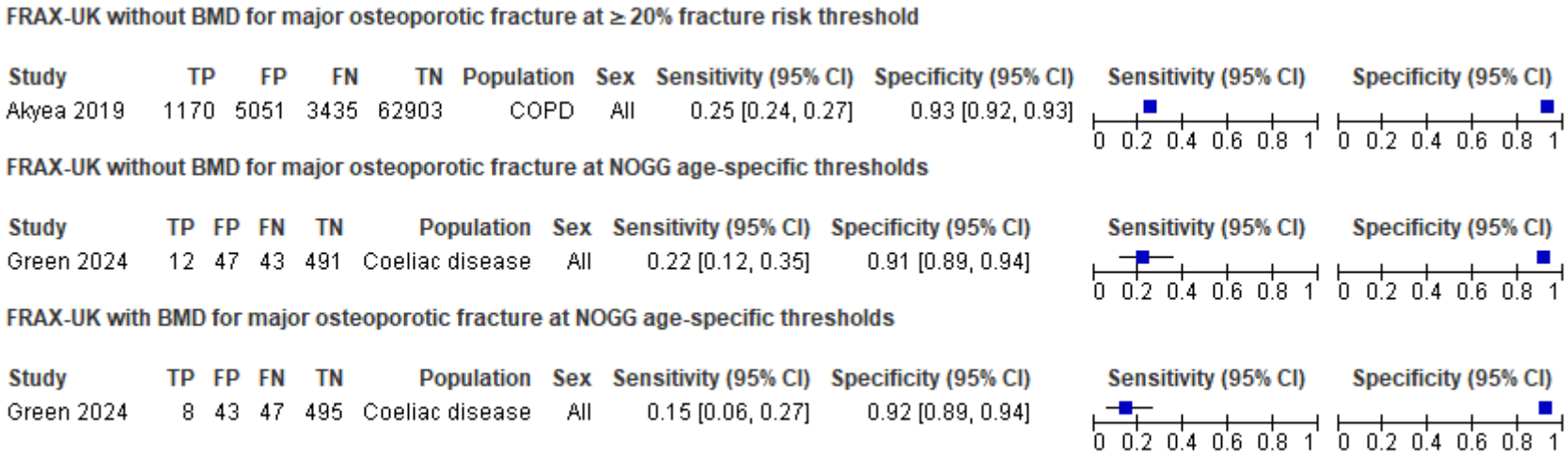
Overall Risk of bias and Applicability	Risk of bias	High *(QFracture 2012: High RoB for predictors (diagnosis of disease status likely to vary across participants) and outcome (composite fracture outcome) domains.)*
Overall Risk of bias and Applicability	Directness	Partially applicable *(includes participants not at suspected risk or not at risk of fragility fracture; women < 65 and men < 75 years)*

Appendix E Forest plots

E.1 Accuracy of risk assessment tools for predicting the risk of fragility fractures

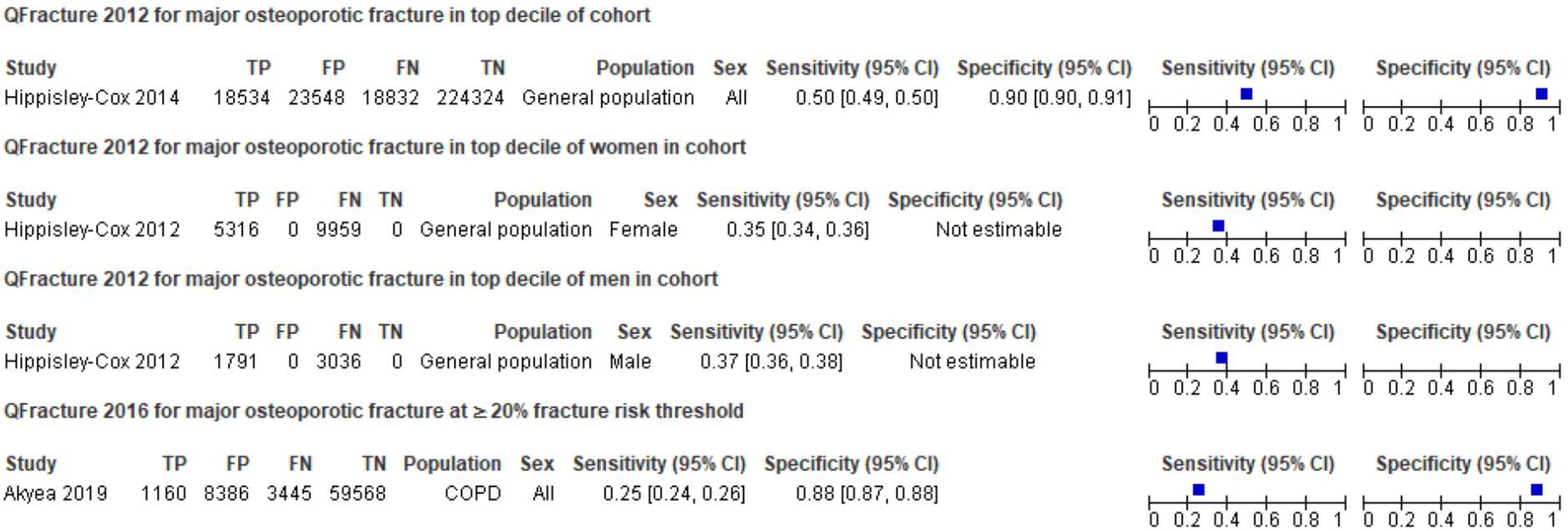
E.1.1 Summary of sensitivity and specificity data

Figure 2: Predictive accuracy of FRAX-UK for 10-year risk of major osteoporotic fracture in high-fracture risk populations



Note: Raw data calculated using the RevMan calculator and the reported sensitivity or specificity, number of people experiencing fractures and number of people above stated threshold.

Figure 3: Predictive accuracy of QFracture for 10-year risk of major osteoporotic fracture in general or high-fracture risk populations



Note: Raw data for all studies calculated using the RevMan calculator and the reported sensitivity or specificity, number of people experiencing fractures and number of people above stated threshold. For Hippisley-Cox 2012, specificity was not reported so zeros were entered into the false positive (FP) and true negative (TN) columns.

Figure 4: Predictive accuracy of QFracture for 10-year risk of hip fracture in general or high-fracture risk populations

QFracture 2012 for hip fracture in top decile of cohort

Study	TP	FP	FN	TN	Population	Sex	Sensitivity (95% CI)	Specificity (95% CI)
Hippisley-Cox 2014	17830	24809	8982	233617	General population	All	0.67 [0.66, 0.67]	0.90 [0.90, 0.91]

QFracture 2012 for hip fracture in top decile of women in cohort

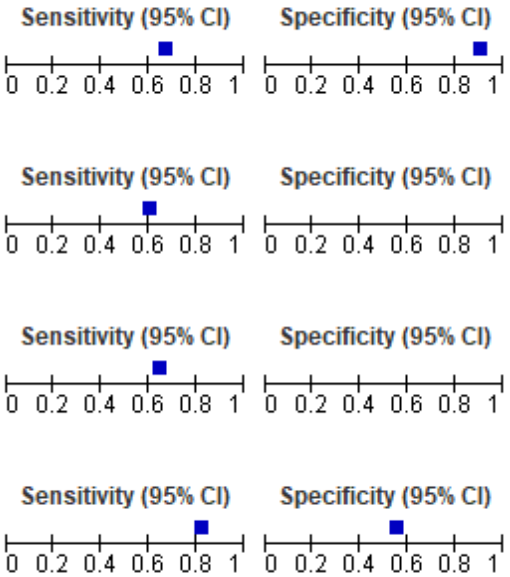
Study	TP	FP	FN	TN	Population	Sex	Sensitivity (95% CI)	Specificity (95% CI)
Hippisley-Cox 2012	3294	0	2215	0	General population	Female	0.60 [0.58, 0.61]	Not estimable

QFracture 2012 for hip fracture in top decile of men in cohort

Study	TP	FP	FN	TN	Population	Sex	Sensitivity (95% CI)	Specificity (95% CI)
Hippisley-Cox 2012	1118	0	621	0	General population	Male	0.64 [0.62, 0.67]	Not estimable

QFracture 2016 for hip fracture ≥ 3% fracture risk threshold

Study	TP	FP	FN	TN	Population	Sex	Sensitivity (95% CI)	Specificity (95% CI)
Akyea 2019	1186	31879	258	39236	COPD	All	0.82 [0.80, 0.84]	0.55 [0.55, 0.56]



Note: Raw data for all studies calculated using the RevMan calculator and the reported sensitivity or specificity, number of people experiencing fractures and number of people above stated threshold. For Hippisley-Cox 2012, specificity was not reported so zeros were entered into the false positive (FP) and true negative (TN) columns.

Appendix F Risk of bias of fragility fracture risk prediction tools (PROBAST)

Table 19: Risk of bias for prediction of 10-year risk of major osteoporotic fracture

Study	Risk of bias				Directness			Overall risk of bias	Overall directness
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome		
Akyea 2019 **FRAX-UK QFracture 2016**	High	High	High	High	Low	Low	Low	High	Directly applicable
Collins 2011 **QFracture 2009**	High	High	High	High	High	Low	Low	High	Partially applicable
Green 2024 **FRAX-UK**	High	High	Low	High	Low	Low	Low	High	Directly applicable
Hippisley-Cox 2009 **FRAX-UK QFracture 2009**	High	High	High	High	High	Low	Low	High	Partially applicable
Hippisley-Cox 2012 **QFracture 2012**	Low	High	High	Low	High	Low	Low	High	Partially applicable
Hippisley-Cox 2014 **QFracture 2012**	Unclear	High	High	Low	High	Low	Low	High	Partially applicable
Ihama 2021 **FRAX-UK QFracture 2016**	Unclear	Low	High	High	Low	Low	Low	High	Directly applicable
Klop 2016 **FRAX-UK**	Low	Low	High	High	Low	Low	Low	High	Directly applicable
Livingstone 2022 **QFracture 2012**	Low	High	High	Low	High	Low	Low	High	Partially applicable

Livingstone 2023 **CFracture**	Low	High	High	Low	High	Low	Low	High	Partially applicable

Table 20: Risk of bias for prediction of 10-year risk of hip fracture

Study	Risk of bias				Directness			Overall risk of bias	Overall directness
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome		
Akyea 2019 **FRAX-UK QFracture 2016**	High	High	High	High	Low	Low	Low	High	Directly applicable
Collins 2011 **QFracture 2009**	High	High	High	High	High	Low	Low	High	Partially applicable
Green 2024 **FRAX-UK**	High	High	Low	High	Low	Low	Low	High	Directly applicable
Hippisley-Cox 2009 **FRAX-UK QFracture 2009**	High	High	High	High	High	Low	Low	High	Partially applicable
Hippisley-Cox 2012 **QFracture 2012**	Low	High	High	Low	High	Low	Low	High	Partially applicable
Hippisley-Cox 2014 **QFracture 2012**	Unclear	High	High	Low	High	Low	Low	High	Partially applicable
Ihama 2021 **FRAX-UK QFracture 2016**	Unclear	Low	High	High	Low	Low	Low	High	Directly applicable
Klop 2016 **FRAX-UK**	Low	Low	High	High	Low	Low	Low	High	Directly applicable

Livingstone 2022 **QFracture 2012**	Low	High	High	Low	High	Low	Low	High	Partially applicable
Livingstone 2023 **CFracture**	Low	High	High	Low	High	Low	Low	High	Partially applicable

Appendix G Excluded studies

G.1 Accuracy of risk assessment tools for predicting the risk of fragility fractures

G.1.1 Clinical studies

Table 21: Studies excluded from the clinical review

Study	Code [Reason]
Adami, Giovanni, Arioli, Giovanni, Bianchi, Gerolamo et al. (2020) Radiofrequency echographic multi spectrometry for the prediction of incident fragility fractures: A 5-year follow-up study. Bone 134: 115297	- Included in Evidence review D
Adami, G, Biffi, A, Porcu, G et al. (2023) A systematic review on the performance of fracture risk assessment tools: FRAX, DeFRA, FRA-HS. Journal of endocrinological investigation	- Systematic review used as source of primary studies *No additional studies identified*
Agarwal, Arnav and Leslie, William D (2022) Fracture prediction tools in diabetes. Current opinion in endocrinology, diabetes, and obesity 29(4): 326-332	- Review article but not a systematic review
Allon, Raviv, Levy, Yahav, Lavi, Idit et al. (2018) How to Best Predict Fragility Fractures: An Update and Systematic Review. The Israel Medical Association journal : IMAJ 20(12): 773-779	- Systematic review used as source of primary studies *No relevant articles*
Anonymous (2023) Correction: Effect of competing mortality risks on predictive performance of the QFracture risk prediction tool for major osteoporotic fracture and hip fracture: external validation cohort study in a UK primary care population. BMJ medicine 2(1): e000316corr1	- Study design not relevant to this review protocol *Published correction of Livingstone 2022*
Ayres, Lachlan Richard Owen, Clarke, Shane, Digby-Bell, Jonathan et al. (2012) Fragility fracture risk in cirrhosis: a comparison of the fracture risk assessment tool, British Society of Gastroenterology and National Institute for Health and Clinical Excellence guidelines. Frontline gastroenterology 3(4): 220-227	- Study does not contain an intervention relevant to this review protocol *Study compares numbers of patients recommended for treatment according to NICE, FRAX-NOGG and BSG guidelines, rather than accuracy of risk prediction tools to predict fracture*

Study	Code [Reason]
Azagra, Rafael, Roca, Genis, Encabo, Gloria et al. (2012) FRAX R tool, the WHO algorithm to predict osteoporotic fractures: the first analysis of its discriminative and predictive ability in the Spanish FRIDEX cohort. BMC musculoskeletal disorders 13: 204	- Study does not contain an intervention relevant to this review protocol *Non-UK FRAX study*
Azagra, Rafael, Zwart, Marta, Encabo, Gloria et al. (2016) Rationale of the Spanish FRAX model in decision-making for predicting osteoporotic fractures: an update of FRIDEX cohort of Spanish women. BMC musculoskeletal disorders 17: 262	- Included in evidence review D
Bach-Mortensen, P, Hyldstrup, L, Appleyard, M et al. (2006) Digital x-ray radiogrammetry identifies women at risk of osteoporotic fracture: results from a prospective study. Calcified tissue international 79(1): 1-6	- Data not reported in an extractable format or a format that can be analysed *No relevant outcomes reported*
Barda, Noam, Yona, Gal, Rothblum, Guy N et al. (2021) Addressing bias in prediction models by improving subpopulation calibration. Journal of the American Medical Informatics Association : JAMIA 28(3): 549-558	- Study does not contain an intervention relevant to this review protocol
Barr, R J, Adebajo, A, Fraser, W D et al. (2005) Can peripheral DXA measurements be used to predict fractures in elderly women living in the community?. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 16(10): 1177-83	- Study does not contain an intervention relevant to this review protocol *Peripheral DXA of the heel*
Battaglino, R., Cobb, G., Nguyen, N. et al. (2018) The discriminative ability of fraxto identify prevalent post-SCI lower extremity osteoporotic fractures. Annals of Physical and Rehabilitation Medicine	- Conference abstract
Bauer, D C, Ewing, S K, Cauley, J A et al. (2007) Quantitative ultrasound predicts hip and non-spine fracture in men: the MrOS study. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 18(6): 771-7	- Population not relevant to this review protocol *Mean age of men<75 years*
Beaudoin, C, Moore, L, Gagne, M et al. (2019) Performance of predictive tools to identify	- Systematic review used as source of primary studies

Study	Code [Reason]
individuals at risk of non-traumatic fracture: a systematic review, meta-analysis, and meta-regression. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 30(4): 721-740	*No relevant articles*
Biamonte, Federica, Pepe, Jessica, Colangelo, Luciano et al. (2024) Assessment of trabecular bone score (TBS) in the prediction of vertebral fracture in postmenopausal osteoporosis. Bone 190: 117307	- Study design not relevant to this review protocol *Case control study*
Bioletto, Fabio, Barale, Marco, Maiorino, Federica et al. (2024) Trabecular Bone Score as a Marker of Skeletal Fragility Across the Spectrum of Chronic Kidney Disease: A Systematic Review and Meta-analysis. The Journal of clinical endocrinology and metabolism 109(7): e1534-e1543	- Systematic review used as source of primary studies *No additional studies identified*
Biver, Emmanuel, Durosier-Izart, Claire, Chevalley, Thierry et al. (2018) Evaluation of Radius Microstructure and Areal Bone Mineral Density Improves Fracture Prediction in Postmenopausal Women. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 33(2): 328-337	- Data not reported in an extractable format or a format that can be analysed *aBMD + age reported only*
Black, Dennis M, Bouxsein, Mary L, Marshall, Lynn M et al. (2008) Proximal femoral structure and the prediction of hip fracture in men: a large prospective study using QCT. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 23(8): 1326-33	- Population not relevant to this review protocol *Mean age of men<75 years*
Boehm, Elena, Kraft, Eduard, Biebl, Johanna Theresia et al. (2024) Quantitative computed tomography has higher sensitivity detecting critical bone mineral density compared to dual-energy X-ray absorptiometry in postmenopausal women and elderly men with osteoporotic fractures: a real-life study. Archives of orthopaedic and trauma surgery 144(1): 179-188	- Study design not relevant to this review protocol *Retrospective study*
Bolland, Mark J, Siu, Amanda Ty, Mason, Barbara H et al. (2011) Evaluation of the FRAX and Garvan fracture risk calculators in older women. Journal of bone and mineral research:	- Included in evidence review D

Study	Code [Reason]
the official journal of the American Society for Bone and Mineral Research 26(2): 420-7	
Bonaccorsi, Gloria, Fila, Enrica, Messina, Carmelo et al. (2017) Comparison of trabecular bone score and hip structural analysis with FRAX R in postmenopausal women with type 2 diabetes mellitus. Aging clinical and experimental research 29(5): 951-957	- Study design not relevant to this review protocol *Case control study*
Briot, Karine and Roux, Christian (2005) What is the role of DXA, QUS and bone markers in fracture prediction, treatment allocation and monitoring?. Best practice & research. Clinical rheumatology 19(6): 951-64	- Review article but not a systematic review
Brismar, Torkel B; Janszky, Imre; Toft, L I M (2010) Calcaneal BMD Obtained by Dual X-Ray and Laser Predicts Future Hip Fractures-A Prospective Study on 4 398 Swedish Women. Journal of osteoporosis 2010: 875647	- Data not reported in an extractable format or a format that can be analysed *Calcaneal (heel) BMD study*
Boutroy, S, Hans, D, Sornay-Rendu, E et al. (2013) Trabecular bone score improves fracture risk prediction in non-osteoporotic women: the OFELY study. Osteoporosis international: a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 24(1): 77-85	- Included in evidence review D
Briot, Karine, Paternotte, Simon, Kolta, Sami et al. (2013) FRAX R: prediction of major osteoporotic fractures in women from the general population: the OPUS study. PloS one 8(12): e83436	- Included in evidence review D
Butscheidt, Sebastian, Rolvien, Tim, Vettorazzi, Eik et al. (2018) Trabecular bone microarchitecture predicts fragility fractures in postmenopausal women on denosumab treatment. Bone 114: 246-251	- Study design not relevant to this review protocol *Retrospective study*
Byberg, Liisa, Gedeborg, Rolf, Cars, Thomas et al. (2012) Prediction of fracture risk in men: a cohort study. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 27(4): 797-807	- Study does not contain an intervention relevant to this review protocol *Non-UK risk prediction model study*
Campillo-Sanchez, F., Usategui-Martin, R., Gil, J. et al. (2021) Fracture risk predictors of a	- Study design not relevant to this review protocol

Study	Code [Reason]
postmenopausal female population by binary statistical procedure CART. Revista de Osteoporosis y Metabolismo Mineral 12(4): 122-128	*Retrospective study*
Catalano, Antonino, Morabito, Nancy, Basile, Giorgio et al. (2013) Fracture risk assessment in postmenopausal women referred to an Italian center for osteoporosis: a single day experience in Messina. Clinical cases in mineral and bone metabolism : the official journal of the Italian Society of Osteoporosis, Mineral Metabolism, and Skeletal Diseases 10(3): 191-4	- Study design not relevant to this review protocol *Association study, no relevant outcomes*
Center, Jacqueline R, Nguyen, Tuan V, Pocock, Nick A et al. (2004) Volumetric bone density at the femoral neck as a common measure of hip fracture risk for men and women. The Journal of clinical endocrinology and metabolism 89(6): 2776-82	- Included in evidence review D
Cepollaro, C, Gonnelli, S, Pondrelli, C et al. (1997) The combined use of ultrasound and densitometry in the prediction of vertebral fracture. The British journal of radiology 70(835): 691-6	- Study design not relevant to this review protocol *Cross-sectional study*
Champakanath, Anagha, Keshawarz, Amena, Pyle, Laura et al. (2021) Fracture risk assessment (FRAX) without BMD and risk of major osteoporotic fractures in adults with type 1 diabetes. Bone 143: 115614	- Study does not contain an intervention relevant to this review protocol *Non-UK FRAX study*
Chan, Mei Y, Nguyen, Nguyen D, Center, Jacqueline R et al. (2012) Absolute fracture-risk prediction by a combination of calcaneal quantitative ultrasound and bone mineral density. Calcified tissue international 90(2): 128-36	- Secondary publication of an included study that does not provide any additional relevant information *DOES cohort study 1994-2009. Study excluded because more recent study on same measurement method/measure (DXA BMD-FN) from this cohort is included in review (Chan 2013, which covers 1994-2011).*
Chan, M Y, Nguyen, N D, Center, J R et al. (2013) Quantitative ultrasound and fracture risk prediction in non-osteoporotic men and women as defined by WHO criteria. Osteoporosis international: a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 24(3): 1015-22	- Included in evidence review D

Study	Code [Reason]
Chapurlat, Roland, Bui, Minh, Sornay-Rendu, Elisabeth et al. (2020) Deterioration of Cortical and Trabecular Microstructure Identifies Women With Osteopenia or Normal Bone Mineral Density at Imminent and Long-Term Risk for Fragility Fracture: A Prospective Study. Journal of bone and mineral research: the official journal of the American Society for Bone and Mineral Research 35(5): 833-844	- Included in evidence review D
Chen, J S, March, L M, Cumming, R G et al. (2009) Role of quantitative ultrasound to predict fracture among institutionalised older people with a history of fracture. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 20(1): 105-12	- Data not reported in an extractable format or a format that can be analysed *QUS study, no relevant outcomes*
Chen, Lin, Wu, Xin-Yi, Jin, Qi et al. (2023) The correlation between osteoporotic vertebrae fracture risk and bone mineral density measured by quantitative computed tomography and dual energy X-ray absorptiometry: a systematic review and meta-analysis. European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society 32(11): 3875-3884	- Comparator in study does not match that specified in this review protocol *Compares QCT and DXA*
Chen, Peiqi, Krege, John H, Adachi, Jonathan D et al. (2009) Vertebral fracture status and the World Health Organization risk factors for predicting osteoporotic fracture risk. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 24(3): 495-502	- Study does not contain an intervention relevant to this review protocol *Non-UK validated risk prediction model study*
Chen, Sy-Jou, Chen, Yi-Ju, Cheng, Chui-Hsuan et al. (2016) Comparisons of Different Screening Tools for Identifying Fracture/Osteoporosis Risk Among Community-Dwelling Older People. Medicine 95(20): e3415	- Reference standard in study does not match that specified in protocol *Uses injurious falls (unintentional loss of balance with body hitting floor or ground from standing height or less resulting in any outpatient or emergency room visit or hospital admission) as reference standard*
Cheneymann, Andia, Therkildsen, Josephine, Rasmussen, Laust Dupont et al. (2024) Developing Cut-off Values for Low and Very Low Bone Mineral Density at the Thoracic Spine Using Quantitative Computed Tomography. Calcified tissue international 115(4): 421-431	- Study design not relevant to this review protocol *Cross-sectional study*

Study	Code [Reason]
Cheng, Xiaoguang, Wang, Ling, Wang, Qianqian et al. (2014) Validation of quantitative computed tomography-derived areal bone mineral density with dual energy X-ray absorptiometry in an elderly Chinese population. Chinese medical journal 127(8): 1445-9	- Study design not relevant to this review protocol *QCT association study, no relevant outcomes*
Cheung, Wing-Hoi, Hung, Vivian Wing-Yin, Cheuk, Ka-Yee et al. (2021) Best Performance Parameters of HR-pQCT to Predict Fragility Fracture: Systematic Review and Meta-Analysis. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 36(12): 2381-2398	- Systematic review used as source of primary studies *No additional studies identified*
Collinge, Cory A, Lebus, George, Gardner, Michael J et al. (2010) A comparison of quantitative ultrasound of the calcaneus with dual-energy x-ray absorptiometry in hospitalized orthopaedic trauma patients. Journal of orthopaedic trauma 24(3): 176-80	- Study design not relevant to this review protocol *Cross-sectional study*
Collins, Gary S and Michaelsson, Karl (2012) Fracture risk assessment: state of the art, methodologically unsound, or poorly reported?. Current osteoporosis reports 10(3): 199-207	- Review article but not a systematic review
Crandall, C.J., Larson, J., Wright, N.C. et al. (2020) Serial Bone Density Measurement and Incident Fracture Risk Discrimination in Postmenopausal Women. JAMA Internal Medicine 180(9): 1232-1240	Crandall, C.J., Larson, J., Wright, N.C. et al. (2020) Serial Bone Density Measurement and Incident Fracture Risk Discrimination in Postmenopausal Women. JAMA Internal Medicine 180(9): 1232-1240
Crandall, C.J., Schousboe, J.T., Morin, S.N. et al. (2019) Performance of FRAX and FRAX-Based Treatment Thresholds in Women Aged 40 and Older: The Manitoba BMD Registry. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research	- Study does not contain an intervention relevant to this review protocol *Non-UK FRAX validation study*
Cummings, S R, Marcus, R, Palermo, L et al. (1994) Does estimating volumetric bone density of the femoral neck improve the prediction of hip fracture? A prospective study. Study of Osteoporotic Fractures Research Group. Journal of bone and mineral research: the official journal of the American Society for Bone and Mineral Research 9(9): 1429-32	- Included in Evidence review D
Cummins, N M, Poku, E K, Towler, M R et al. (2011) clinical risk factors for osteoporosis in	- Study design not relevant to this review protocol

Study	Code [Reason]
Ireland and the UK: a comparison of FRAX and QFractureScores. Calcified tissue international 89(2): 172-7	*Case-control study*
D'Amore, Simona, Sano, Hiroshige, Chappell, Daniel David George et al. (2023) Radiographic Cortical Thickness Index Predicts Fragility Fracture in Gaucher Disease. Radiology 307(1): e212779	- Study design not relevant to this review protocol *Retrospective study*
Dagan, Noa, Cohen-Stavi, Chandra, Leventer-Roberts, Maya et al. (2017) External validation and comparison of three prediction tools for risk of osteoporotic fractures using data from population based electronic health records: retrospective cohort study. BMJ (Clinical research ed.) 356: i6755	- Study does not contain an intervention relevant to this review protocol *Non-UK FRAX and QFracture study*
Damilakis, John, Papadokostakis, George, Perisinakis, Kostas et al. (2007) Hip fracture discrimination by the Achilles Insight QUS imaging device. European journal of radiology 63(1): 59-62	- Study design not relevant to this review protocol *Case-control study*
Dargent-Molina, P, Piault, S, Breart, G et al. (2003) A comparison of different screening strategies to identify elderly women at high risk of hip fracture: results from the EPIDOS prospective study. Osteoporosis international: a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 14(12): 969-77	- Included in Evidence review D
Dargent-Molina, P, Schott, A M, Hans, D et al. (1999) Separate and combined value of bone mass and gait speed measurements in screening for hip fracture risk: results from the EPIDOS study. Epidemiologie de l'Osteoporose. Osteoporosis international: a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 9(2): 188-92	- Included in Evidence review D
Desbiens, Louis-Charles; Goupil, Remi; Mac-Way, Fabrice (2020) Predictive value of quantitative ultrasound parameters in individuals with chronic kidney disease: A population-based analysis of CARTaGENE. Bone 130: 115120	- Population not relevant to this review protocol *Mean age of women<65 years; mean age of men <75 years*
Desbiens, Louis-Charles, Sidibe, Aboubacar, Beaudoin, Claudia et al. (2020) Comparison of	- Study does not contain an intervention relevant to this review protocol

Study	Code [Reason]
Fracture Prediction Tools in Individuals Without and With Early Chronic Kidney Disease: A Population-Based Analysis of CARTaGENE. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 35(6): 1048-1057	*Non-UK FRAX and QFracture study*
Dhiman, Paula, Andersen, Stig, Vestergaard, Peter et al. (2018) Does bone mineral density improve the predictive accuracy of fracture risk assessment? A prospective cohort study in Northern Denmark. BMJ open 8(4): e018898	- Study does not contain an intervention relevant to this review protocol *Non-UK validated risk prediction model*
Diacinti, D, Pisani, D, Barone-Adesi, F et al. (2010) A new predictive index for vertebral fractures: the sum of the anterior vertebral body heights. Bone 46(3): 768-73	- Data not reported in an extractable format or a format that can be analysed *Outcome is prediction of vertebral fracture only*
Donaldson, Meghan G, Palermo, Lisa, Schousboe, John T et al. (2009) FRAX and risk of vertebral fractures: the fracture intervention trial. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 24(11): 1793-9	- Data not reported in an extractable format or a format that can be analysed *Vertebral fractures reported only*
Durosier, C, Hans, D, Krieg, M A et al. (2008) Defining risk thresholds for a 10-year probability of hip fracture model that combines clinical risk factors and quantitative ultrasound: results using the EPISEM cohort. Journal of clinical densitometry : the official journal of the International Society for Clinical Densitometry 11(3): 397-403	- Comparator in study does not match that specified in this review protocol *Reference standard not described*
Ensrud, Kristine E, Lui, Li-Yung, Taylor, Brent C et al. (2009) A comparison of prediction models for fractures in older women: is more better?. Archives of internal medicine 169(22): 2087-94	- Study does not contain an intervention relevant to this review protocol *BMD +age reported only*
Ensrud, Kristine E, Schousboe, John T, Crandall, Carolyn J et al. (2024) Hip Fracture Risk Assessment Tools for Adults Aged 80 Years and Older. JAMA network open 7(6): e2418612	- Included in Evidence review D
Fidler, Jeff L, Murthy, Naveen S, Khosla, Sundeep et al. (2016) Comprehensive Assessment of Osteoporosis and Bone Fragility with CT Colonography. Radiology 278(1): 172-80	- Study design not relevant to this review protocol *Retrospective cohort study*

Study	Code [Reason]
FitzGerald, Gordon, Compston, Juliet E, Chapurlat, Roland D et al. (2014) Empirically based composite fracture prediction model from the Global Longitudinal Study of Osteoporosis in Postmenopausal Women (GLOW). The Journal of clinical endocrinology and metabolism 99(3): 817-26	- Study does not contain an intervention relevant to this review protocol *Validation of non-UK risk prediction model*
Friis-Holmberg, Teresa, Rubin, Katrine Hass, Brixen, Kim et al. (2014) Fracture risk prediction using phalangeal bone mineral density or FRAX(R)?-A Danish cohort study on men and women. Journal of clinical densitometry : the official journal of the International Society for Clinical Densitometry 17(1): 7-15	- Study does not contain an intervention relevant to this review protocol *DXA of phalanges (fingers/toes) not relevant bone assessment technique*
Frost, M L; Blake, G M; Fogelman, I (2002) A comparison of fracture discrimination using calcaneal quantitative ultrasound and dual X-ray absorptiometry in women with a history of fracture at sites other than the spine and hip. Calcified tissue international 71(3): 207-11	- Study design not relevant to this review protocol *Cross-sectional diagnostic study*
Fu, Y, Li, C, Luo, W et al. (2021) Fragility fracture discriminative ability of radius quantitative ultrasound: a systematic review and meta-analysis. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 32(1): 23-38	- Systematic review used as source of primary studies *No additional studies identified*
Gates, Michelle, Pillay, Jennifer, Nuspl, Megan et al. (2023) Screening for the primary prevention of fragility fractures among adults aged 40 years and older in primary care: systematic reviews of the effects and acceptability of screening and treatment, and the accuracy of risk prediction tools. Systematic reviews 12(1): 51	- Systematic review used as source of primary studies *Canadian review of FRAX and other models, UK studies not included in review*
Gnudi, S; Ripamonti, C; Malavolta, N (2000) Quantitative ultrasound and bone densitometry to evaluate the risk of nonspine fractures: a prospective study. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 11(6): 518-23	- No relevant outcome reported *Non-spine fracture only reported*
Gnudi, S; Sitta, E; Fiumi, N (2007) Bone density and geometry in assessing hip fracture risk in	- Study design not relevant to this review protocol

Study	Code [Reason]
post-menopausal women. The British journal of radiology 80(959): 893-7	
Gong, Joanna Y, Chiang, Cherie, Wark, John D et al. (2024) Bone Density and Trabecular Bone Score Decline Rapidly in the First Year After Bone Marrow Transplantation with a Marked Increase in 10-Year Fracture Risk. Calcified tissue international 114(4): 377-385	- Study design not relevant to this review protocol *Retrospective non-UK FRAX and TBS study*
Goodhand, J R, Kamperidis, N, Nguyen, H et al. (2011) Application of the WHO fracture risk assessment tool (FRAX) to predict need for DEXA scanning and treatment in patients with inflammatory bowel disease at risk of osteoporosis. Alimentary pharmacology & therapeutics 33(5): 551-8	- Comparator in study does not match that specified in this review protocol *Reference standard is WHO BMD cutoff definitions of osteoporosis/osteopenia*
Gourlay, Margaret L, Ritter, Victor S, Fine, Jason P et al. (2017) Comparison of fracture risk assessment tools in older men without prior hip or spine fracture: the MrOS study. Archives of osteoporosis 12(1): 91	- Population not relevant to this review protocol *Mean age of men<75 years*
Grampp, S, Genant, H K, Mathur, A et al. (1997) Comparisons of noninvasive bone mineral measurements in assessing age-related loss, fracture discrimination, and diagnostic classification. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 12(5): 697-711	- Study design not relevant to this review protocol *Case control study*
Grassi, Lorenzo, Vaananen, Sami P, Jehpsson, Lars et al. (2023) 3D Finite Element Models Reconstructed From 2D Dual-Energy X-Ray Absorptiometry (DXA) Images Improve Hip Fracture Prediction Compared to Areal BMD in Osteoporotic Fractures in Men (MrOS) Sweden Cohort. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 38(9): 1258-1267	- Population not relevant to this review protocol *Population-based MrOS cohort*
Guo, SiJia, An, Ning, Lin, JiSheng et al. (2022) Comparison of four tools to identify painful new osteoporotic vertebral fractures in the postmenopausal population in Beijing. Frontiers in endocrinology 13: 1013755	- Study does not contain an intervention relevant to this review protocol *Non-UK FRAX and BMD retrospective validation study*
Gupta, Ayushman, Greening, Neil J, Evans, Rachael A et al. (2019) Prospective risk of osteoporotic fractures in patients with advanced	- Study design not relevant to this review protocol *Cross-sectional study design*

Study	Code [Reason]
chronic obstructive pulmonary disease. Chronic respiratory disease 16: 1479972318769763	
Guthrie, Bruce, Rogers, Gabriel, Livingstone, Shona et al. (2024) The implications of competing risks and direct treatment disutility in cardiovascular disease and osteoporotic fracture: risk prediction and cost effectiveness analysis. Health and social care delivery research 12(4): 1-275	- Secondary publication of an included study that does not provide any additional relevant information *Performance data for external validation of QFracture and internal validation of CFracture reported in Livingstone 2023*
Hadji, P, Hars, O, Gorke, K et al. (2000) Quantitative ultrasound of the os calcis in postmenopausal women with spine and hip fracture. Journal of clinical densitometry : the official journal of the International Society for Clinical Densitometry 3(3): 233-9	- Study design not relevant to this review protocol *Cross-sectional study design*
Hans, D, Allaoua, S, Genton, L et al. (2002) Is time since hip fracture influencing the discrimination between fractured and nonfractured subjects as assessed at the calcaneum by three technologically different quantitative ultrasound devices?. Calcified tissue international 71(6): 485-92	- Study design not relevant to this review protocol *Case-control study*
Hans, D, Schott, A M, Duboeuf, F et al. (2004) Does follow-up duration influence the ultrasound and DXA prediction of hip fracture? The EPIDOS prospective study. Bone 35(2): 357-63	- Included in Evidence review D
Hansen, K E, Blank, R D, Palermo, L et al. (2014) What analytic method should clinicians use to derive spine T-scores and predict incident fractures in men? Results from the MrOS study. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 25(9): 2181-8	- Study design not relevant to this review protocol *Case-control study*
Henry, Margaret J, Pasco, Julie A, Merriman, Elizabeth N et al. (2011) Fracture risk score and absolute risk of fracture. Radiology 259(2): 495-501	- Study does not contain an intervention relevant to this review protocol *Non-UK validated risk prediction tool (FRISK)*
Hillier, Teresa A, Stone, Katie L, Bauer, Doug C et al. (2007) Evaluating the value of repeat bone mineral density measurement and prediction of fractures in older women: the study of osteoporotic fractures. Archives of internal medicine 167(2): 155-60	- Included in Evidence review D

Study	Code [Reason]
Hollaender, R, Hartl, F, Krieg, M-A et al. (2009) Prospective evaluation of risk of vertebral fractures using quantitative ultrasound measurements and bone mineral density in a population-based sample of postmenopausal women: results of the Basel Osteoporosis Study. Annals of the rheumatic diseases 68(3): 391-6	- Data not reported in an extractable format or a format that can be analysed *Vertebral fractures reported only*
Holloway-Kew, Kara L, Betson, Amelia G, Anderson, Kara B et al. (2024) Associations between ultra-distal forearm bone mineral density and incident fracture in women. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 35(6): 1019-1027	- Population not relevant to this review protocol *Mean age by sex not reported. Median age of fracture patients 70.1; median age of no fracture patients 61.5*
Hsieh, Wen-Tung, Groot, Tom Maarten, Yen, Hung-Kuan et al. (2024) Validation of Ten Osteoporosis Screening Tools in Rural Communities of Taiwan. Calcified tissue international 115(5): 507-515	- Study does not contain an intervention relevant to this review protocol *Non-UK FRAX study*
Hu, Jia-Sen, Jin, Ya-Ping, Wu, Ji-Kui et al. (2024) Skeletal muscle index based on CT at the 12th thoracic spine level can predict osteoporosis and fracture risk: a propensity score-matched cohort study. Frontiers in medicine 11: 1387807	- Study design not relevant to this review protocol *Retrospective cohort study*
Huopio, J, Kroger, H, Honkanen, R et al. (2004) Calcaneal ultrasound predicts early postmenopausal fractures as well as axial BMD. A prospective study of 422 women. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 15(3): 190-5	- No relevant outcome reported *Fracture outcome includes non-fragility fractures*
Iki, M, Winzenrieth, R, Tamaki, J et al. (2021) Predictive ability of novel volumetric and geometric indices derived from dual-energy X-ray absorptiometric images of the proximal femur for hip fracture compared with conventional areal bone mineral density: the Japanese Population-based Osteoporosis (JPOS) Cohort Study. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 32(11): 2289-2299	- Population not relevant to this review protocol *Mean age of women<65 years*

Study	Code [Reason]
Ishii, Shinya, Greendale, Gail A, Cauley, Jane A et al. (2012) Fracture risk assessment without race/ethnicity information. The Journal of clinical endocrinology and metabolism 97(10): 3593-602	- Population not relevant to this review protocol *Mean age of women<65 years*
Jacobs, J.W.G., Da Silva, J.A.P., Armbrecht, G. et al. (2010) Prediction of vertebral fractures is specific for gender and site of bone mineral density measurement. Journal of Rheumatology 37(1): 149-154	- Study does not contain an intervention relevant to this review protocol *Study looked into accuracy of BMD measurement at predicting the presence of vertebral deformities*
Jamal, S A, Cheung, A M, West, S L et al. (2012) Bone mineral density by DXA and HR pQCT can discriminate fracture status in men and women with stages 3 to 5 chronic kidney disease. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 23(12): 2805-13	- Study design not relevant to this review protocol *Cross-sectional study*
Jamal, S A; West, S L; Nickolas, T L (2014) The clinical utility of FRAX to discriminate fracture status in men and women with chronic kidney disease. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 25(1): 71-6	- Study design not relevant to this review protocol *Cross-sectional study*
James, Herbert 3rd, Aleksic, Ilija, Bienz, Marc Nicolas et al. (2014) Comparison of fracture risk assessment tool score to bone mineral density for estimating fracture risk in patients with advanced prostate cancer on androgen deprivation therapy. Urology 84(1): 164-8	- Study does not contain an intervention relevant to this review protocol *Non-UK FRAX*
Jazinizadeh, F; Adachi, J D; Quenneville, C E (2020) Advanced 2D image processing technique to predict hip fracture risk in an older population based on single DXA scans. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 31(10): 1925-1933	- Study does not contain an intervention relevant to this review protocol *Machine-learning/statistical shape model study*
Jiang, X., Gruner, M., Tremollieres, F. et al. (2015) Diagnostic accuracy of FRAX in predicting the 10-year risk of osteoporotic	- Conference abstract

Study	Code [Reason]
fractures: A systematic review and meta-analysis. Menopause: 1392	
Johanen, Astera, Jonasson, Grethe, Lund, Henrik et al. (2021) Trabecular bone patterns as a fracture risk predictor: a systematic review. Acta odontologica Scandinavica 79(7): 482-491	- Systematic review used as source of primary studies *No additional studies identified*
Johansson, H, Kanis, J A, Oden, A et al. (2014) Impact of femoral neck and lumbar spine BMD discordances on FRAX probabilities in women: a meta-analysis of international cohorts. Calcified tissue international 95(5): 428-35	- Study design not relevant to this review protocol *Meta-analysis of international cohort examining potential discordance in BMD measurements and FRAX*
Jonasson, Grethe B, Sundh, Valter, Hakeberg, Magnus et al. (2018) Evaluation of clinical and radiographic indices as predictors of osteoporotic fractures: a 10-year longitudinal study. Oral surgery, oral medicine, oral pathology, and oral radiology 125(5): 487-494	- Study does not contain an intervention relevant to this review protocol *Non-UK FRAX study*
Kanis, J A, Johansson, H, Harvey, N C et al. (2021) An assessment of intervention thresholds for very high fracture risk applied to the NOGG guidelines : A report for the National Osteoporosis Guideline Group (NOGG). Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 32(10): 1951-1960	- Systematic review used as source of primary studies *No additional studies identified*
Kanis, J A, Johnell, O, Oden, A et al. (2008) FRAX and the assessment of fracture probability in men and women from the UK. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 19(4): 385-97	- Data not reported in an extractable format or a format that can be analysed *No relevant outcomes*
Kanis, John A, Harvey, Nicholas C, Cooper, Cyrus et al. (2016) A systematic review of intervention thresholds based on FRAX : A report prepared for the National Osteoporosis Guideline Group and the International Osteoporosis Foundation. Archives of osteoporosis 11(1): 25	- Systematic review used as source of primary studies *No additional studies identified*
Kim, Dam, Cho, Soo-Kyung, Kim, Ji Young et al. (2016) Association between trabecular bone score and risk factors for fractures in Korean	- Data not reported in an extractable format or a format that can be analysed

Study	Code [Reason]
female patients with rheumatoid arthritis. Modern rheumatology 26(4): 540-5	*AUC data reported only for vertebral fractures*
Kim, Hyoun-Ah, Lee, Hyun Young, Jung, Ju-Yang et al. (2020) Trabecular Bone Score Is a Useful Parameter for the Prediction of Vertebral Fractures in Patients With Polymyalgia Rheumatica. Journal of clinical densitometry : the official journal of the International Society for Clinical Densitometry 23(3): 373-380	- Study design not relevant to this review protocol *Cross-sectional diagnostic study*
Krieg, Marc-Antoine, Cornuz, Jacques, Ruffieux, Christiane et al. (2006) Prediction of hip fracture risk by quantitative ultrasound in more than 7000 Swiss women > or =70 years of age: comparison of three technologically different bone ultrasound devices in the SEMOF study. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 21(9): 1457-63	- Included in Evidence review D
Kuzma, Martin, Hans, Didier, Koller, Tomas et al. (2018) Less strict intervention thresholds for the FRAX and TBS-adjusted FRAX predict clinical fractures in osteopenic postmenopausal women with no prior fractures. Journal of bone and mineral metabolism 36(5): 580-588	- Included in Evidence review D
Kwok, T, Khoo, C C, Leung, J et al. (2012) Predictive values of calcaneal quantitative ultrasound and dual energy X ray absorptiometry for non-vertebral fracture in older men: results from the MrOS study (Hong Kong). Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 23(3): 1001-6	- Population not relevant to this review protocol *Population-based MrOs cohort, community-dwelling older men, not selected on basis of fracture risk*
Lamy, O., Krieg, M.-A., Stoll, D. et al. (2012) The OsteoLaus Cohort Study: Bone mineral density, micro-architecture score and vertebral fracture assessment extracted from a single DXA device in combination with clinical risk factors improve significantly the identification of women at high risk of fracture. Osteologie 21(2): 77-82	- Data not reported in an extractable format or a format that can be analysed *BMD adjusted for age and BMI reported only*
Leonhardt, Yannik, May, Pauline, Gordijenko, Olga et al. (2020) Opportunistic QCT Bone Mineral Density Measurements Predicting Osteoporotic Fractures: A Use Case in a Prospective Clinical Cohort. Frontiers in endocrinology 11: 586352	- Included in Evidence review D

Study	Code [Reason]
Leslie, W.D., Anderson, W.A., Metge, C.J. et al. (2007) Clinical risk factors for fracture in postmenopausal Canadian women: A population-based prevalence study. Bone 40(4): 991-996	- Study does not contain an intervention relevant to this review protocol
Leslie, W D, Aubry-Rozier, B, Lix, L M et al. (2014) Spine bone texture assessed by trabecular bone score (TBS) predicts osteoporotic fractures in men: the Manitoba Bone Density Program. Bone 67: 10-4	- Included in Evidence review D
Leslie, W.D.; Metge, C.; Ward, L. (2003) Contribution of clinical risk factors to bone density-based absolute fracture risk assessment in postmenopausal women. Osteoporosis International 14(4): 334-338	- Comparator in study does not match that specified in this review protocol *Confirmation of fracture not established through clinical report, self-report, or radiograph*
Leslie, William D, Tsang, James F, Caetano, Patricia A et al. (2007) Effectiveness of bone density measurement for predicting osteoporotic fractures in clinical practice. The Journal of clinical endocrinology and metabolism 92(1): 77-81	- Included in Evidence review D
Li, Guowei, Leslie, William D, Kovacs, Christopher S et al. (2020) Combining Frailty and Trabecular Bone Score Did Not Improve Predictive Accuracy in Risk of Major Osteoporotic Fractures. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 35(6): 1058-1064	- Study does not contain an intervention relevant to this review protocol *TBS and FRAX-Canada study, no BMD reported*
Lin, Wentao, He, Chaoqin, Xie, Faqin et al. (2023) Quantitative CT screening improved lumbar BMD evaluation in older patients compared to dual-energy X-ray absorptiometry. BMC geriatrics 23(1): 231	- Data not reported in an extractable format or a format that can be analysed *vertebral fractures reported*
Lin, Wentao, He, Chaoqin, Xie, Faqin et al. (2023) Discordance in lumbar bone mineral density measurements by quantitative computed tomography and dual-energy X-ray absorptiometry in postmenopausal women: a prospective comparative study. The spine journal : official journal of the North American Spine Society 23(2): 295-304	- Study design not relevant to this review protocol *Diagnostic study design*
Link, Thomas M, Vieth, Volker, Matheis, Julia et al. (2002) Bone structure of the distal radius and the calcaneus vs BMD of the spine and proximal	- Study design not relevant to this review protocol

Study	Code [Reason]
femur in the prediction of osteoporotic spine fractures. European radiology 12(2): 401-8	*Case-control study*
Liu, Hau, Paige, Neil M, Goldzweig, Caroline L et al. (2008) Screening for osteoporosis in men: a systematic review for an American College of Physicians guideline. Annals of internal medicine 148(9): 685-701	- Systematic review used as source of primary studies *No additional studies identified*
Liu, Zhenyu, Gao, Hua, Bai, Xiaodong et al. (2017) Evaluation of Singh Index and Osteoporosis Self-Assessment Tool for Asians as risk assessment tools of hip fracture in patients with type 2 diabetes mellitus. Journal of orthopaedic surgery and research 12(1): 37	- Study design not relevant to this review protocol *Retrospective study*
Loffler, Maximilian T, Jacob, Alina, Scharr, Andreas et al. (2021) Automatic opportunistic osteoporosis screening in routine CT: improved prediction of patients with prevalent vertebral fractures compared to DXA. European radiology 31(8): 6069-6077	- Study design not relevant to this review protocol *Retrospective study*
Long, Yujia; Leslie, William D; Luo, Yunhua (2015) Study of DXA-derived lateral-medial cortical bone thickness in assessing hip fracture risk. Bone reports 2: 44-51	- Study design not relevant to this review protocol *Case-control study*
Lopez, Ben, Meertens, Robert, Gundry, Mike et al. (2024) A comparison between IBEX bone health applied to digital radiographs and dual-energy X-ray absorptiometry at the distal-third and ultra-distal regions of the radius. BMC musculoskeletal disorders 25(1): 575	- Comparator in study does not match that specified in this review protocol *Reference standard not relevant (DXA)*
Lott, Ariana, Pflug, Emily M, Parola, Rown et al. (2022) Predicting the Subsequent Contralateral Hip Fracture: Is FRAX the Answer?. Journal of orthopaedic trauma 36(12): 599-603	- Study does not contain an intervention relevant to this review protocol *Not FRAX-UK study*
Mackey, Dawn C, Eby, Jean Gaare, Harris, Fran et al. (2007) Prediction of clinical non-spine fractures in older black and white men and women with volumetric BMD of the spine and areal BMD of the hip: the Health, Aging, and Body Composition Study*. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 22(12): 1862-8	- No relevant outcome reported *Reports AUC data for non-spine fracture only*
Mariotti, Veronica, Page, David B, Davydov, Oksana et al. (2017) Assessing fracture risk in	- Study design not relevant to this review protocol

Study	Code [Reason]
early stage breast cancer patients treated with aromatase-inhibitors: An enhanced screening approach incorporating trabecular bone score. Journal of bone oncology 7: 32-37	*Retrospective non-UK FRAX study*
Marques, Andrea, Ferreira, Ricardo J O, Santos, Eduardo et al. (2015) The accuracy of osteoporotic fracture risk prediction tools: a systematic review and meta-analysis. Annals of the rheumatic diseases 74(11): 1958-67	- Systematic review used as source of primary studies *No additional studies identified*
Marques, Andrea, Lucas, Raquel, Simoes, Eugenia et al. (2017) Do we need bone mineral density to estimate osteoporotic fracture risk? A 10-year prospective multicentre validation study. RMD open 3(2): e000509	- Population not relevant to this review protocol *Mean age by sex not reported. Mean age of cohort 58.2 years*
Mazziotti, Gherardo, Vena, Walter, Pedersini, Rebecca et al. (2022) Prediction of vertebral fractures in cancer patients undergoing hormone deprivation therapies: Reliability of who fracture risk assessment tool (frax) and bone mineral density in real-life clinical practice. Journal of bone oncology 33: 100421	- Study does not contain an intervention relevant to this review protocol *Non-UK FRAX Study (Italy)*
McCloskey, E V, Harvey, N C, Johansson, H et al. (2022) Fracture risk assessment by the FRAX model. Climacteric : the journal of the International Menopause Society 25(1): 22-28	- Review article but not a systematic review
McCloskey, E V, Kanis, J A, Oden, A et al. (2015) Predictive ability of heel quantitative ultrasound for incident fractures: an individual-level meta-analysis. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 26(7): 1979-87	- Data not reported in an extractable format or a format that can be analysed *No relevant outcomes reported*
Michalski, A S, Besler, B A, Burt, L A et al. (2021) Opportunistic CT screening predicts individuals at risk of major osteoporotic fracture. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 32(8): 1639-1649	- Study design not relevant to this review protocol *Retrospective study*
Mikolajewicz, Nicholas, Bishop, Nick, Burghardt, Andrew J et al. (2020) HR-pQCT Measures of Bone Microarchitecture Predict Fracture: Systematic Review and Meta-Analysis. Journal of bone and mineral research : the official	- Systematic review used as source of primary studies *No additional studies identified*

Study	Code [Reason]
journal of the American Society for Bone and Mineral Research 35(3): 446-459	
Moayyeri, Alireza, Kaptoge, Stephen, Dalzell, Nichola et al. (2009) Is QUS or DXA better for predicting the 10-year absolute risk of fracture?. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 24(7): 1319-25	- Population not relevant to this review protocol *Mean age by sex not reported. Mean age of cohort 60.5 years*
Moayyeri, Alireza, Kaptoge, Stephen, Dalzell, Nichola et al. (2009) The effect of including quantitative heel ultrasound in models for estimation of 10-year absolute risk of fracture. Bone 45(2): 180-4	- Data not reported in an extractable format or a format that can be analysed
Nayak, S, Edwards, D L, Saleh, A A et al. (2014) Performance of risk assessment instruments for predicting osteoporotic fracture risk: a systematic review. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 25(1): 23-49	- Systematic review used as source of primary studies *No additional studies identified*
Nguyen, N D, Frost, S A, Center, J R et al. (2008) Development of prognostic nomograms for individualizing 5-year and 10-year fracture risks. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 19(10): 1431-44	- Secondary publication of an included study that does not provide any additional relevant information *DOES cohort study 1989-2004. Study excluded because more recent study on same measurement method/measure (DXA BMD-FN) from this cohort is included in review (Chan 2013, which covers 1994-2011).*
Nguyen, Tuan V; Center, Jacqueline R; Eisman, John A (2013) Individualized fracture risk assessment: progresses and challenges. Current opinion in rheumatology 25(4): 532-41	- Review article but not a systematic review
Pinheiro, M M, Castro, C H M, Frisoli, A Jr et al. (2003) Discriminatory ability of quantitative ultrasound measurements is similar to dual-energy X-ray absorptiometry in a Brazilian women population with osteoporotic fracture. Calcified tissue international 73(6): 555-64	- Study design not relevant to this review protocol *Cross-sectional study*
Pisani, Paola, Conversano, Francesco, Muratore, Maurizio et al. (2023) Fragility Score: a REMS-based indicator for the prediction of incident fragility fractures at 5 years. Aging clinical and experimental research	- Population not relevant to this review protocol *Population not selected by risk of fragility fracture/indication for scan*

Study	Code [Reason]
Popp, A W, Senn, C, Franta, O et al. (2009) Tibial or hip BMD predict clinical fracture risk equally well: results from a prospective study in 700 elderly Swiss women. Osteoporosis international: a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 20(8): 1393-9	- Included in Evidence review D
Prince, Richard, Khoo, Benjamin, Brown, Keenan et al. (2023) Differences in Femoral Neck and Trochanteric Structure in Elderly Women Prior to Hip Fracture: Role in Hip Fracture Prediction. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research	- Data not reported in an extractable format or a format that can be analysed *aBMD + age reported only*
Prins, S H, Jorgensen, H L, Jorgensen, L V et al. (1998) The role of quantitative ultrasound in the assessment of bone: a review. Clinical physiology (Oxford, England) 18(1): 3-17	- Systematic review used as source of primary studies *No additional studies identified*
Pulkkinen, Pasi, Partanen, Juha, Jalovaara, Pekka et al. (2004) Combination of bone mineral density and upper femur geometry improves the prediction of hip fracture. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 15(4): 274-80	- Study design not relevant to this review protocol *Case control study*
Rampersad, C, Whitlock, R H, Leslie, W D et al. (2020) Trabecular bone score in patients with chronic kidney disease. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 31(10): 1905-1912	- Study design not relevant to this review protocol *Retrospective cohort study*
Robbins, John, Aragaki, Aaron K, Kooperberg, Charles et al. (2007) Factors associated with 5-year risk of hip fracture in postmenopausal women. JAMA 298(20): 2389-98	- Population not relevant to this review protocol *Mean age of women <65 years*
Rubin, Katrine Hass, Friis-Holmberg, Teresa, Hermann, Anne Pernille et al. (2013) Risk assessment tools to identify women with increased risk of osteoporotic fracture: complexity or simplicity? A systematic review. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 28(8): 1701-17	- Systematic review used as source of primary studies *No additional studies identified*

Study	Code [Reason]
Schacter, G Isanne and Leslie, William D (2017) DXA-Based Measurements in Diabetes: Can They Predict Fracture Risk?. Calcified tissue international 100(2): 150-164	- Study design not relevant to this review protocol *Review article*
Schacter, G I, Leslie, W D, Majumdar, S R et al. (2017B) Clinical performance of an updated trabecular bone score (TBS) algorithm in men and women: the Manitoba BMD cohort. Osteoporosis international: a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 28(11): 3199-3203	- Included in Evidence review D
Schousboe, John T, Vo, Tien, Taylor, Brent C et al. (2016) Prediction of Incident Major Osteoporotic and Hip Fractures by Trabecular Bone Score (TBS) and Prevalent Radiographic Vertebral Fracture in Older Men. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 31(3): 690-7	- Study does not contain an intervention relevant to this review protocol *Non-UK FRAX and non-UK validated risk prediction model study*
Serrano-Montalban, Beatriz, Arias, Angel, Friginal-Ruiz, Ana Belen et al. (2017) The Use of the Fracture Risk Assessment (FRAX R) Tool in Predicting Risk of Fractures in Patients With Inflammatory Bowel Disease: A Systematic Review. Journal of clinical densitometry : the official journal of the International Society for Clinical Densitometry 20(2): 180-187	- Systematic review used as source of primary studies *No additional studies identified*
Shahla, Ahmad (2011) Validity of bone mineral density and WHO fracture risk assessment thresholds in hip fractures. Archives of Iranian medicine 14(5): 352-4	- Study design not relevant to this review protocol *Cross-sectional study*
Sharma, Ashish, Sinha, Rahul Janak, Singh, Vishwajeet et al. (2019) Implications of the Fracture Risk Assessment Algorithm for the assessment and improvement of bone health in patients with prostate cancer: A comprehensive review. Turkish journal of urology 45(4): 245-253	- Systematic review used as source of primary studies *No additional studies identified*
Sharma, Seema and Khandelwal, Sunila (2010) Effective risk assessment tools for osteoporosis in the Indian menopausal female. Journal of mid-life health 1(2): 79-85	- Study does not contain an intervention relevant to this review protocol

Study	Code [Reason]
Sheu, Yahtyng, Zmuda, Joseph M, Boudreau, Robert M et al. (2011) Bone strength measured by peripheral quantitative computed tomography and the risk of nonvertebral fractures: the osteoporotic fractures in men (MrOS) study. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 26(1): 63-71	- No relevant outcome reported *AUC data reported for non-spine fracture only*
Shevroja, Enisa, Reginster, Jean-Yves, Lamy, Olivier et al. (2023) Update on the clinical use of trabecular bone score (TBS) in the management of osteoporosis: results of an expert group meeting organized by the European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases (ESCEO), and the International Osteoporosis Foundation (IOF) under the auspices of WHO Collaborating Center for Epidemiology of Musculoskeletal Health and Aging. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 34(9): 1501-1529	- Systematic review used as source of primary studies *No additional studies identified*
Stewart, A, Felsenberg, D, Eastell, R et al. (2006) Relationship between risk factors and QUS in a European Population: The OPUS study. Bone 39(3): 609-15	- Study design not relevant to this review protocol *Prevalence study from survey data, no relevant intervention*
Stewart, Alison; Kumar, Vinod; Reid, David M (2006) Long-term fracture prediction by DXA and QUS: a 10-year prospective study. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 21(3): 413-8	- Population not relevant to this review protocol *Mean age by sex not reported. Mean age of cohort 47 years*
Sun, Xuemei, Chen, Yancong, Gao, Yinyan et al. (2022) Prediction Models for Osteoporotic Fractures Risk: A Systematic Review and Critical Appraisal. Aging and disease 13(4): 1215-1238	- Systematic review used as source of primary studies *No additional studies identified*
Szulc, Pawel; Boutroy, Stephanie; Chapurlat, Roland (2018) Prediction of Fractures in Men Using Bone Microarchitectural Parameters Assessed by High-Resolution Peripheral Quantitative Computed Tomography-The Prospective STRAMBO Study. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 33(8): 1470-1479	- Population not relevant to this review protocol *Mean age by sex not reported. Mean age of cohort 72.1 years*

Study	Code [Reason]
Tanaka, Shiro, Yoshimura, Noriko, Kuroda, Tatsuhiko et al. (2010) The Fracture and Immobilization Score (FRISC) for risk assessment of osteoporotic fracture and immobilization in postmenopausal women--A joint analysis of the Nagano, Miyama, and Taiji Cohorts. Bone 47(6): 1064-70	- Population not relevant to this review protocol *Mean age by sex not reported. Mean age of cohort 59.5 years*
Tei, Randi M H, Plana-Ripoll, Oleguer, Brink, Ole et al. (2019) An Optimised Fracture Liaison Service Model: Maintained Diagnostic Sensitivity Despite Reduced Number of Diagnostic Tests Performed. Calcified tissue international 104(6): 641-649	- Study does not contain an intervention relevant to this review protocol *'Interventions' are clinical risk factors*
Testi, D, Cappello, A, Chiari, L et al. (2001) Comparison of logistic and Bayesian classifiers for evaluating the risk of femoral neck fracture in osteoporotic patients. Medical & biological engineering & computing 39(6): 633-7	- Study design not relevant to this review protocol *Case control study*
Tremollieres, Florence A, Pouilles, Jean-Michel, Drewniak, Nicolas et al. (2010) Fracture risk prediction using BMD and clinical risk factors in early postmenopausal women: sensitivity of the WHO FRAX tool. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 25(5): 1002-9	- Population not relevant to this review protocol *Mean age of women <65 years*
Trimpou, Penelope, Bosaeus, Ingvar, Bengtsson, Bengt-Ake et al. (2010) High correlation between quantitative ultrasound and DXA during 7 years of follow-up. European journal of radiology 73(2): 360-4	- Comparator in study does not match that specified in this review protocol *Reference standard is t-score as assessed by DXA-BMD*
Viswanathan, Meera, Reddy, Shivani, Berkman, Nancy et al. (2018) Screening to Prevent Osteoporotic Fractures: An Evidence Review for the U.S. Preventive Services Task Force.	- Systematic review used as source of primary studies *No additional studies identified*
Viswanathan, Meera, Reddy, Shivani, Berkman, Nancy et al. (2018) Screening to Prevent Osteoporotic Fractures: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA 319(24): 2532-2551	- Systematic review used as source of primary studies *No additional studies identified*
Vogrig, E, Della Martina, M, Xodo, S et al. (2014) Identification of patients with high osteoporosis risk: analysis of FRAX and phalangeal ultrasonography in a female	- Study design not relevant to this review protocol *Non-UK FRAX retrospective cohort study*

Study	Code [Reason]
population in North-East Italy. Minerva ginecologica 66(5): 447-53	
Wu, Q, Magnus, J H, Rice, J C et al. (2010) Does using lower limit of normal values enhance the ability of a single bone mineral density measure to predict fractures?. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 21(11): 1881-8	- Data not reported in an extractable format or a format that can be analysed *Diagnostic data presented in figure*
Yamada, M, Ito, M, Hayashi, K et al. (1994) Dual energy X-ray absorptiometry of the calcaneus: comparison with other techniques to assess bone density and value in predicting risk of spine fracture. AJR. American journal of roentgenology 163(6): 1435-40	- Study design not relevant to this review protocol *Case-control study*
Yamamoto, M, Yamaguchi, T, Yamauchi, M et al. (2007) Bone mineral density is not sensitive enough to assess the risk of vertebral fractures in type 2 diabetic women. Calcified tissue international 80(6): 353-8	- Study design not relevant to this review protocol *Cross sectional study*
Yin, Michael T and Falutz, Julian (2016) How to predict the risk of fracture in HIV?. Current opinion in HIV and AIDS 11(3): 261-7	- Review article but not a systematic review
Zaia, Annamaria, Rossi, Roberto, Galeazzi, Roberta et al. (2021) Fractal lacunarity of trabecular bone in vertebral MRI to predict osteoporotic fracture risk in over-fifties women. The LOTO study. BMC musculoskeletal disorders 22(1): 108	- Data not reported in an extractable format or a format that can be analysed *vertebral fractures reported*
Zarzour, Fatima and Leslie, William D (2024) Fracture Risk Associated with Different Numbers and Combinations of Lumbar Vertebrae: The Manitoba BMD Registry. Journal of clinical densitometry : the official journal of the International Society for Clinical Densitometry 27(3): 101502	- Included in Evidence review D
Zhang, Bo, Zhou, Lu-Ping, Zhang, Xian-Liang et al. (2023) Which Indicator Among Lumbar Vertebral Hounsfield Unit, Vertebral Bone Quality, or Dual-Energy X-Ray Absorptiometry-Measured Bone Mineral Density Is More Efficacious in Predicting Thoracolumbar Fragility Fractures?. Neurospine 20(4): 1193-1204	- Data not reported in an extractable format or a format that can be analysed *vertebral fractures reported*

Study	Code [Reason]
Zoccarato, Francesca, Ceolin, Chiara, Trevisan, Caterina et al. (2022) Comparison between real-world practice and application of the FRAX algorithm in the treatment of osteoporosis. Aging clinical and experimental research 34(11): 2807-2814	- Study design not relevant to this review protocol *Retrospective cohort study*

1

Appendix H Research Recommendation

H.1 What is the validity of CFracture risk prediction tool for predicting the risk of fragility fractures in adults, including those who have had a previous fragility fracture?

H.1.1 Why this is important

Risk prediction tools play an important role in identify which people are at risk of fragility fractures and should have imaging. The performance of the recommended UK-validated tools in terms of both calibration and discrimination was acceptable for the estimation of an individual's 10-year risk of major osteoporotic fractures. However, more accurate prediction of fragility fracture in adults at risk or suspected risk of fracture could improve identification of people who might benefit from early intervention and reduce fracture incidence.

Importance to 'patients' or the population	More accurate tools for prediction of fragility fracture in adults could improve the identification of people at risk of fractures. Subsequent treatment where appropriate would reduce fractures and associated morbidity.
Relevance to NICE guidance	High. The research is essential to inform future updates as more robust and accurate tools should in principle be used for risk prediction. Studies comparing accuracy and calibration of new tool to those currently in use in same population at risk or suspected risk of fragility fracture needed to directly compare performance of tools.
Relevance to the NHS	The NHS and commissioners of services would need to consider how to encourage uptake and use of new risk prediction tool.
National priorities	Not a national priority. High relevance to the NICE guideline for Osteoporosis.
Current evidence base	Although the current evidence is limited to one study, calculator for CFracture is not currently available to public. The data and model is available.
Equality considerations	None.

H.1.2 Modified PICOTS table

Population	Adults at suspected risk of fragility fractures (people with or at risk of primary or secondary osteoporosis or have had a previous fragility fracture)
Intervention	- CFracture
Comparator	- QFracture and/or FRAX-UK with or without BMD. Reference standard ideally radiologically confirmed fracture.
Target condition	- 10-year risk of major osteoporotic fracture (hip, clinical spine, shoulder, and forearm/wrist) - 10-year risk of hip fracture
Statistical outcomes	- Overall fit (Cox-Snell R^2, Brier score) - Calibration (plots including smoothed flexible calibration curve with CIs to allow future development studies; calibration sloped, calibration-in-the-large, O:E ratio, calibration index) - Discrimination (AUC/c-statistic, sensitivity, specificity, predictive values) - Reclassification - Clinical utility (net benefit, associated decision curve)

Study design	Prospective or retrospective external validation cohort study preferably comparing all 3 risk prediction tools. Large study required although can be retrospective
Timeframe	Completed prior to future updates of the osteoporosis guideline to inform future recommendations.

Abbreviations: AUC, area under the receiver operating curve; BMD, bone mineral density; O:E, observed:expected.