National Institute for Health and Care Excellence

Draft for consultation

Osteoporosis: risk assessment

[D] Accuracy of bone assessment methods

NICE guideline <number>

Evidence reviews underpinning recommendations 1.4.1-1.4.4 and recommendations 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 All NICE guidance is subject to regular review and may be updated or withdrawn.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. Accuracy of bone assessment methods

1.1. Review questions: What is the accuracy of bone assessment methods for predicting fragility fractures in adults, including those who have had a previous fragility fracture?

1.1.1. Introduction

Various bone assessment methods have been proposed to predict fragility fracture risk, including dual x-ray absorptiometry (DXA), various forms of quantitative computed tomography (QCT), quantitative ultrasound, and more recently radiofrequency echographic multi spectrometry (REMS). This review examines the accuracy of bone assessment methods to predict fragility fracture in adults.

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 osteoporosis or have had a previous fragility fracture).

Intervention	Predictive accuracy of major osteoporotic fracture or hip fracture using:
	• Dual X-ray absorptiometry (DXA) or dual x-ray and laser (DXL) or hip, spine, or forearm
	  ○ Areal bone mineral density (aBMD) only
	  ○ aBMD with trabecular bone score (TBS) assessment
	• Quantitative computed tomography scans (QCT) including asynchronous calibration QCT (phantom-less scanning); high-resolution QCT (HR-QCT); peripheral QCT (pQCT); and photon counting CT
	  ○ Volumetric BMD (vBMD)
	• Quantitative ultrasound (QUS), for example, Bindex
	  ○ Broadband ultrasound attenuation (BUA)
	  ○ Speed of sound (SOS)
	  ○ Quantitative ultrasound index (QUI) or Stiffness Index (SI) (both combine BUA and SOS measurements)
	• Digital radiography (IBEX BH software)
	  ○ aBMD
	Bone assessment methods do not require validation in UK population.

Outcomes	• Discrimination: c-statistic/AUC, for overall discrimination • Sensitivity/specificity, predictive values at specific threshold
Target condition	Fragility fracture • Major osteoporotic fracture (hip, clinical vertebral, humerus, forearm) • Hip fracture Assessed using the reference standard: combination of clinical review, self-report, and confirmation of fracture by radiography.
Setting	• Any setting
Study design	• Prospective cohort studies

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 below.

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

The target population for this review is adults 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). As such, studies were included if the participants were referred for DXA or presented to hospital with fragility fractures. Studies in which the cohort was a general unselected population were included if they were high risk as defined by the previous NICE Osteoporosis guideline CG146 (published 2012), that is, if the mean age of the participants was ≥65 years for women or ≥75 years for men. If a study did not report age by sex, then it was excluded.

Since age is a significant predictor of fragility fracture, age-adjusted BMD scores were preferred to both unadjusted scores and scores adjusted for other risk factors when extracting data. Meta-analysis of AUC data was conducted where appropriate using the package *metamisc* version 0.1.8 in R and the valmeta function, which performs a random-effects meta-analysis on studies in line with the methods described in Debray 2018. Meta-analysis was conducted using both a frequentist (restricted maximum likelihood ratio) and a Bayesian (rjags) model.

Heterogeneity was assessed using I^2 (the proportion of total variance on the logit scale) and visual inspection of forest plots. Additional tests for heterogeneity were also considered alongside these, including τ^2 (tau-squared: the amount or magnitude of between-study variance on logit scale) and the 95% prediction intervals (the range which the results of future studies are likely to be within). I^2 is a relative measure of the proportion of total variance that is due to between-study heterogeneity compared to within-study variance. By contrast, τ^2 is an absolute measure on the logit scale of between-study variance. When studies are precise (that is, the standard errors of the AUC are small and the within-study variance is therefore small), a small τ^2 can lead to a high I^2 because the within-study variance is smaller than the between-study variance. Meta regression was used to explore heterogeneity due to follow up time and treatment history. Prespecified subgroup analysis by treatment status (on treatment versus not on treatment) to explain heterogeneity was not possible as only a single study included participants not on treatment whilst the remaining studies included participants who were on treatment, mixed populations or did not report the treatment status.

1.1.4. Bone assessment methods to predict fragility fracture

1.1.4.1. Dual X-ray absorptiometry

The most widely used method to assess bone health is DXA, which is a low dose ionising radiation technology. Areal BMD is typically measured at different sites of the skeleton, including those most vulnerable to fracture, such as the lumbar spine and the proximal femur.

1.1.4.2. Quantitative computed tomography

Standard quantitative computed tomography (QCT) can be performed on any CT scanner using an external bone mineral calibration phantom, which is placed underneath the patient usually in a supine position, and appropriate analysis software. Cross calibration allowing comparison of BMD measurements from different phantoms is needed. Volumetric or 3D QCT (vQCT) is generally preferred to the original single-slice (2D) QCT because it has

increased precision and is easier to perform. However, radiation dose with vQCT can be substantially higher than single-slice QCT.

1.1.4.3. Quantitative ultrasound

Quantitative ultrasound (QUS) is a non-ionising portable ultrasound technology. It has been used to measure bone mineral health via the speed at which the ultrasound signal travels through bone (speed of sound, SOS) and how much its amplitude is attenuated (broadband ultrasound attenuation, BUA). Several composite measures combining SOS and BUA have also been proposed (for example, the Stiffness Index).

Most QUS devices are designed for use at one skeletal site, but multisite devices are becoming available. The technical characteristics of QUS devices such as frequency of emitted ultrasounds, transmission bone pathways, and QUS variables can differ and results between machines are difficult to compare.

Radiofrequency Echographic Multi Spectrometry (REMS) is a recent portable non-ionising, rapid ultrasound technology that can measure BMD and bone quality at the femoral neck or lumbar spine. Automated identification of bone parts using image processing of raw, unfiltered native ultrasound signals acquired during an echographic scan provides quantitative measurements (for example: BMD, T- and Z-scores) and qualitative assessment (for example: Fragility Score) of bone health.

1.1.4.4. Digital radiography

Digital radiography uses digital technology and dedicated software analysis to capture and store radiographs, allowing estimation of areal BMD (aBMD) via comparison of simulation and real X-ray images.

1.1.5. Bone assessment evidence

Evidence was identified regarding the predictive accuracy of bone assessment methods for predicting fragility fractures in adults including those who have had a previous fragility fracture. The bone assessment methods and the specific outcomes are summarised in Table 2. Full details can be found in Appendix D, Appendix E, and Appendix F.

Evidence related to effectiveness of bone assessment techniques was sought as part of Evidence review E.

1.1.5.1. Included studies

Twenty-three studies examined the accuracy of bone assessment methods in adults at risk or suspected risk of fragility fracture to discriminate between adults who will develop a fracture and those who will not. Mean follow up time in the studies ranged from 2.1 years to greater than 10 years, with over half the studies following patients for over 5 years.

Population

Six studies were in adults, men or women referred for DXA (Adami 2020, Azagra 2016, Leslie 2007a, Leslie 2014, Schacter 2017b, Zarzour 2024). One study was in adults hospitalised for fragility fracture (Leonhardt 2020).

Fifteen studies were in population-based cohorts of postmenopausal women unselected based on fracture risk or suspected risk with a mean age ≥ 65.0 years (Bolland 2011, Boutroy 2013, Briot 2013, Center 2004, Chan 2013, Chapurlat 2020, Crandall 2020, Cummings 1994, Dargent-Molina 1999, Dargent-Molina 2003, Hans 2004, Hillier 2007, Krieg 2006, Kuzma 2018, Popp 2009). One study was in unselected adults ≥ 80 years (Ensrud 2024).

Treatment history

Three studies were in adults who had not received anti-osteoporosis treatment (including hormone replacement therapy) that might affect bone density (Azagra 2016, Bolland 2011) or reported separate data for this subgroup (Briot 2013). Nine studies included some adults who were on treatment at baseline or during follow-up (Boutroy 2013, Briot 2013, Crandall 2020, Ensrud 2024, Hillier 2007, Krieg 2006, Kuzma 2018, Leslie 2014, Popp 2009). The remaining studies did not report use of treatment at baseline or follow up.

Interventions and reference standard

Most studies examined the accuracy of BMD as measured by DXA to predict major osteoporotic or hip fracture using aBMD measurements of the femoral neck (-FN) or lumbar spine (-LS) in postmenopausal or older women. Six studies examined the accuracy of BMD as measured by DXA at the total hip (-TH). Four studies examined the accuracy of QUS measures (for example, BUA, SOS) to predict fragility fracture in postmenopausal women (Dargent-Molina 1999/2003, Hans 2004, Krieg 2006). One study examined the accuracy of the ultrasound technology REMS compared to DXA-based BMD measurements to predict fracture risk in postmenopausal women (Adami 2020). One study examined the accuracy of vBMD as measured by QCT to predict major osteoporotic or hip fracture in a mixed population hospitalised for fragility fracture (Leonhardt 2020). No studies examining the accuracy of digital radiography to predict fragility fracture were identified.

The reference standard for most studies (that is, fracture ascertainment) was self-reported incident fractures during follow up typically confirmed by clinical review and radiographic evidence (that is radiographs and reports). Four studies used diagnostic codes from healthcare records to ascertain fracture (Leslie 2007, Leslie 2014, Schacter 2017B, Zarzour 2024).

Target condition and statistical measures

Fifteen studies reported the outcome of major osteoporotic fracture (Adami 2020, Azagra 2016, Bolland 2011, Boutroy 2013, Briot 2013, Chan 2013, Chapurlat 2020, Crandall 2020, Kuzma 2018, Leonhardt 2020, Leslie 2007a, Leslie 2014, Popp 2009, Schacter 2017b, Zarzour 2024). Most defined it as including (at the least) hip, shoulder, forearm and clinical vertebral (spine) fragility fractures, though some studies were more inclusive in the type of fragility fractures predicted.

Seventeen studies reported the outcome of hip fracture (Adami 2020, Azagra 2016, Bolland 2011, Center 2004, Chan 2013, Crandall 2020, Cummings 1994, Dargent-Molina 2009, Dargent-Molina 2003, Ensrud 2024, Hans 2004, Hillier 2007, Krieg 2006, Leslie 2007a, Leslie 2014, Schacter 2017b, Zarzour 2024).

Most included studies reported AUC data, with only 3 studies reporting sensitivity and/or specificity at a specific fracture risk threshold (Center 2004, Dargent-Molina 1999 and 2003).

For further details on the included studies, see Table 2 for summary tables and Appendix D for full study characteristics.

1.1.5.2. Excluded studies

See the excluded studies list in Appendix G.

1.1.5.3. Studies included in the bone assessment methods to predict fragility fracture evidence review

Table 2: Summary of prospective cohort studies included in the bone assessment methods to predict fragility fracture evidence review

Study Country	Number of participants (Men/Women) and population Cohort	Follow up Age, years (SD)[a] Number on treatment at baseline/during FU	Bone assessment method: measure	Reference standard	Predicted outcomes	Discrimination outcomes
Adami 2020 Italy	• N=1516 Caucasian women referred for DXA, 30-90 years • Single hospital 2013-2016	• Mean FU: 3.7 years (SD 0.8) • Median age: 60.0 (IQR 54-66) • On treatment at baseline: NR • Received treatment during FU: NR	• DXA: BMD-LS • DXA: BMD-FN • REMS: BMD-LS • REMS: BMS-FN	Medical reports based on imaging (for example, radiographs or vertebral morphometry)	• Major osteoporotic fracture[b] • Hip fracture	• AUC adjusted for age, adjusted for age and BMI
Azagra 2016 Spain	• N=1,308 White women referred for DXA, 40-90 years • FRIDEX cohort 2000-2010	• FU: ≥10 years • Median age (IQR): ○ With fracture: 71.0 (62.8-74.3) ○ Without fracture: 59.7 (54.1-64.8) • On treatment at baseline: No	• DXA: BMD-FN	Self-report with fracture confirmed by medical records	• Major osteoporotic fracture • Hip fracture	• AUC

Study Country	Number of participants (Men/Women) and population Cohort	Follow up Age, years (SD)[a] Number on treatment at baseline/during FU	Bone assessment method: measure	Reference standard	Predicted outcomes	Discrimination outcomes
		• Received treatment during FU: NR				
Bolland 2011 New Zealand	• N=1,422 healthy postmenopausal women≥55 years-old • Originally participated in 5-year calcium supplement trial (1998-2005), followed up 2008-2009	• Mean FU: 8.8 years (2.4), range 0.2-11.4 • Mean age: 74.2 (4.2) • On treatment at baseline: No • Received treatment during FU: NR	• DXA: BMD-FN	Self-report with fracture confirmed by radiographs/radiographic review (original 5-year RCT) Self-report (subsequent extension study)	• Major osteoporotic fracture • Hip fracture	• AUC
Boutroy 2013 France	• N=582 postmenopausal women, 31-89 years • Retrospective analysis of OFELY cohort, recruited 1992-1993, BMD measurements 2000-2001, followed up to 2009	• Mean FU: 7.8 (SD 1.3) • Mean age: ○ With fracture: 70.4 (9.4) ○ Without fracture: 65.3 (7.6) • On treatment at baseline: NR • Received treatment during FU: 41%	• DXA: BMD-LS • DXA: BMD-TH	Clinical review confirmed by radiographs	• Major osteoporotic fracture[c]	• AUC

Study Country	Number of participants (Men/Women) and population Cohort	Follow up Age, years (SD)[a] Number on treatment at baseline/during FU	Bone assessment method: measure	Reference standard	Predicted outcomes	Discrimination outcomes
Briot 2013 Europe (France, Germany, UK)	• N=1,748 postmenopausal women≥55 years • OPUS cohort 1999-2001	• Mean FU: 6.04 years (range 4.5-7.5) • Mean age: 66.1 (6.8) • On treatment at baseline or during FU: 60% • Received treatment during FU: 8.9%	• DXA: BMD-FN	Self-report with fracture confirmed by written radiographic or surgical reports	• Major osteoporotic facture	• AUC adjusted for age and fragility fracture history • AUC subgroup treatment • AUC subgroup no treatment
Center 2004 Australia	• N=1584 (658/926) community-dwelling adults≥60 years-old • DOES cohort 1989-2002	• FU: up to 13.0 years • Mean age ○ Women with HF (SD): 77.0 (7.0) ○ Women without HF (SD): 70.0 (7.0) ○ Men with HF (SD): 76.0 (8.0) • Men without HF (SD): 69.0 (6.0)	• DXA: BMD-FN adjusted for age	Review of radiology reports and clinical review	• Hip fracture	• Sensitivity/specificity at -2.5 SD WHO threshold for diagnosing osteoporosis

Study Country	Number of participants (Men/Women) and population Cohort	Follow up Age, years (SD)[a] Number on treatment at baseline/during FU	Bone assessment method: measure	Reference standard	Predicted outcomes	Discrimination outcomes
		• On treatment at baseline: NR • Received treatment during FU: NR				
Chan 2013 Australia	• N=702 (390/312) non-osteoporotic community-dwelling adults≥60 years-old • DOES cohort 1994-2011	• Median FU: 12.0 (range 0.1-17.0) • Mean age men: 72.4 • Mean age women: 70.9 • Age range 62-90 years • On treatment at baseline: NR • Received treatment during FU: NR	• DXA: BMD-FN • QUS-heel: BUA (CUBA sonometer)	Review of radiology reports and clinical review	• Hip fracture in women	• AUC adjusted for age, falls and fracture history
Chapurlat 2020 France	• N=2128 postmenopausal women including ○ N=589 from OFELY cohort 2006-2008 ○ N=1539 from QUALYOR	• OFELY median FU: 9.4 years • QUALYOR FU: 5.0 years • Mean age 67.0 (range 42.0-96.0)	• DXA: BMD-FN	Radiographs, DXA VFA, or clinical reports	• Major osteoporotic fracture	• AUC

Study Country	Number of participants (Men/Women) and population Cohort	Follow up Age, years (SD)[a] Number on treatment at baseline/during FU	Bone assessment method: measure	Reference standard	Predicted outcomes	Discrimination outcomes
	cohort 2010-2013	• On treatment at baseline: NR • Received treatment during FU: NR				
Crandall 2020 USA	• N=7419 postmenopausal women, 50-79 years • 3 centres, WHI Bone Density Sub-study cohort 2005-2010	• FU:12.1 years (SD 3.4) • Mean age: 66.1 (7.2) • On treatment at baseline: 47% • Received treatment during FU: NR	• DXA: BMD-FN • DXA: BMD-LS • DXA: BMD-TH	Self-report with fracture adjudicated using medical records	• Major osteoporotic fracture[d] • Hip fracture	• AUC adjusted for hormone use, clinic, age, race/ethnicity, fracture history, physical activity, and fall history
Cummings 1994 USA	• N=7963 White women ≥65 years-old • SOF cohort 1986-1988 with BMD measurements obtained 1988-1990	• Mean FU: 2.1 years after BMD measurements obtained • Mean age: 73.2 • On treatment at baseline: NR • Received treatment during FU: NR	• DXA: BMD-FN	Self-report with fracture confirmed by radiologist using radiographic reports or preoperative radiographs	• Hip fracture	• AUC

Study Country	Number of participants (Men/Women) and population Cohort	Follow up Age, years (SD)[a] Number on treatment at baseline/during FU	Bone assessment method: measure	Reference standard	Predicted outcomes	Discrimination outcomes
Dargent-Molina 1999 France	• N=5895 postmenopausal women ≥75 years-old • EPIDOS cohort 1992-1994	• FU: 2.75 years (SD 0.7) • Mean age: ○ With HF: 82.6 (4.5) ○ Without HF: 80.4 (3.7) • On treatment at baseline: NR • Received treatment during FU: NR	• DXA: BMD-FN adjusted for age • QUS-heel: BUA (Achilles sonometer)	Self-report	• Hip fracture	• Sensitivity at top 10%/25%/50% fracture risk (90%/75%/50% threshold)
Dargent-Molina 2003 France	• N=5910 postmenopausal women≥75 years-old • EPIDOS cohort 1992-1994	• FU: 3.7 years (SD 0.8) • Mean age: 80.5 (3.8) • On treatment at baseline: NR • Received treatment during FU: NR	• DXA: BMD-FN • QUS-heel: BUA (Achilles sonometer)	Self-report	• Hip fracture	• Sensitivity/specificity at top 5% fracture risk (95% threshold)
Ensrud 2024 USA	• N=8,890 (3984/4906) community-dwelling	• Mean FU: 4.4 (SD 1.2) years • Mean age: ○ Women: 82.6 (2.7)	• DXA: BMD-FN	Self-report with fracture confirmed by radiology reports	• Hip fracture for men • Hip fracture for women	• AUC

Study Country	Number of participants (Men/Women) and population Cohort	Follow up Age, years (SD)[a] Number on treatment at baseline/during FU	Bone assessment method: measure	Reference standard	Predicted outcomes	Discrimination outcomes
	adults≥80 years-old • SOF, Health ABC, MrOS cohorts 1997-2016	○ Men: 82.7 (2.7) • On treatment at baseline: Men 4.4%; Women 10.2% • Received treatment during FU: NR				
Hans 2004 France	• N=5,898 postmenopausal White women≥75 years-old • EPIDOS cohort 1992-1994	• Mean FU: 3.5 years • Mean age: ○ With HF: 82.6 (4.5) ○ Without HF: 80.4 (3.7) • On treatment at baseline: NR • Received treatment during FU: NR	• DXA: BMD-FN • DXA: BMD-TF • QUS-heel: BUA (Achilles sonometer) • QUS-heel: SOS (Achilles sonometer) • QUS-heel: SI (Achilles sonometer)	Self-report with fracture confirmed by rheumatologist using preoperative radiographs and surgical reports	• Hip fracture	• AUC
Hiller 2007 USA	• N=4,124 community-dwelling postmenopausal women≥65 years-old	• Mean FU: 8.0 years (range 6.3-98) • Mean age: 72.0 (4.0)	• DXA: BMD-TH	Self-report with fracture adjudicated by physician from radiology reports	• Hip fracture	• AUC

Study Country	Number of participants (Men/Women) and population Cohort	Follow up Age, years (SD)[a] Number on treatment at baseline/during FU	Bone assessment method: measure	Reference standard	Predicted outcomes	Discrimination outcomes
	• SOF cohort 1989-2004	• On treatment at baseline: NR • Received treatment during FU: 25%				
Krieg 2006 Switzerland	• N=7,062 postmenopausal women≥70 years-old • SEMOF cohort 1997-1999	• Mean FU: 2.9 years (SD 0.8), • Mean age: 75.2 (3.1) • On treatment at baseline: 2.8% • Received treatment during FU: 6%	• QUS-heel: BUA (GE Lunar Achilles+, Sahara sonometers) • QUS-heel: SOS (GE Lunar Achilles+, Sahara sonometers) • QUS-heel: SI (Achilles+, Sahara sonometer)	Self-report with fracture confirmed by medical report	• Hip fracture	• AUC
Kuzma 2018 Slovakia	• N=127 postmenopausal women 50+ years • Single centre, 2009-2015	• Mean FU: 5.2 years (range 2.6-7.0) • Mean age: 66.1 • On treatment at baseline: No	• DXA: BMD-FN • DXA: BMD-LS	Self-report with fracture confirmed by medical record from surgeon or traumatologist	• Major osteoporotic fracture[e]	• AUC

Study Country	Number of participants (Men/Women) and population Cohort	Follow up Age, years (SD)[a] Number on treatment at baseline/during FU	Bone assessment method: measure	Reference standard	Predicted outcomes	Discrimination outcomes
		• Received treatment during FU: 4%				
Leonhardt 2020 Germany	• N=58 (16/42) adults hospitalised due to fragility fracture • Single hospital, 2015-2018	• FU: ≥3 years • Mean age: 72.8 (10.7) • On treatment at baseline: NR • Received treatment during FU: NR	• QCT: vBMD-LS	New imaging findings or clinical follow up	• Major osteoporotic fracture[e]	• AUC
Leslie 2007A Canada	• N=16,505 women≥50 years old referred for DXA • Manitoba Bone Density Program cohort, recruited 1987-2002, BMD measurements 1998-2002	• Mean FU: 3.2 years (SD 1.5) after BMD measurements • Mean age: 65.0 (9.0) • On treatment at baseline: NR • Received treatment during FU: NR	• DXA: BMD-FN • DXA: BMD-LS • DXA: BMD-TH	Diagnostic codes from healthcare records	• Major osteoporotic fracture • Hip fracture	• AUC
Leslie 2014 Canada	• N=3,620 men≥50 years old referred for DXA	• Mean FU: 4.5 years • Mean age: 67.6 (9.8)	• DXA: BMD-FN • DXA: BMD-LS • DXA: BMD-FN + TBS-LS	Diagnostic codes from healthcare records	• Major osteoporotic fracture • Hip fracture	• AUC

Study Country	Number of participants (Men/Women) and population Cohort	Follow up Age, years (SD)[a] Number on treatment at baseline/during FU	Bone assessment method: measure	Reference standard	Predicted outcomes	Discrimination outcomes
	• Manitoba Bone Density Program cohort, recruited 1987-2011	• On treatment at baseline: 21.3% • Received treatment during FU: NR	• DXA: BMD-LS + TBS-LS			
Popp 2009 Switzerland	• N=637 community-dwelling postmenopausal women, 70-80 years • SEMOF cohort 1998-2002	• Mean FU: 2.8 years (SD 0.6) • Mean age: 76.0 (3.0) • On treatment at baseline: 5% of fracture patients, 12% of no fracture patients • Received treatment during FU: NR	• DXA: BMD-FN • DXA: BMD-LS, ICSD T score • DXA: BMD-TH	Self-report with fracture confirmed by questionnaire to family practitioner or hospital in charge	• Major osteoporotic fracture[e]	• AUC adjusted for age and fracture history
Schacter 2017B Canada	• N=52,084 (4,348/47,736) adults≥40 years-old referred for DXA • Manitoba Bone Density Program 1999-2011	• Mean FU for men: 5.0 years • Mean age for men: 64.0 (12.0) • Mean FU for women: 6.0 years • Mean age for women: 63.0 (11.0)	• DXA: BMD-LS	Diagnostic codes from healthcare records	• Major osteoporotic fracture for men • Major osteoporotic fracture for women • Hip fracture for men • Hip fracture for women	• AUC

Study Country	Number of participants (Men/Women) and population Cohort	Follow up Age, years (SD)[a] Number on treatment at baseline/during FU	Bone assessment method: measure	Reference standard	Predicted outcomes	Discrimination outcomes
		• On treatment at baseline: NR • Received treatment during FU: NR				
Zarzour 2024 Canada	• N=39,727 (3,571/36,156) adults≥40 years-old referred for DXA • Manitoba Bone Density Program cohort, recruited	• Mean FU: 8.7 years (5.2) • Mean age: 62.7 (10.5) • On treatment at baseline: NR • Received treatment during FU: NR	• DXA: BMD-FN • DXA: BMD-LS • DXA: BMD-TH	Diagnostic codes from healthcare records	• Major osteoporotic fracture • Hip fracture	• AUC

Abbreviations: AUC, area under the curve; BMD, bone mineral density; BUA, broadband ultrasonic attenuation; DXA, dual x-ray absorptiometry; FN, femoral neck; FU, follow up; HF, hip fracture; LS, lumbar spine; QUS, quantitative ultrasound; REMS, radiofrequency echographic multi spectrometry; TBS, trabecular bone score; TH, total hip; UD, ultradistal; vBMD, volumetric bone mineral density; WHO, World Health Organization.

Notes:

a. age data is mean and SD unless other stated.

b. reports fragility fractures and includes wrist, spine, shoulder, hip, ribs, forearm, ankle, pelvis, and other fractures.

c. includes non-clinical fractures identified by 4-yearly radiographs.

d. includes hip, spine, radius, ulna, wrist, upper arm, and shoulder fractures.

e. study reports osteoporotic fragility fractures so may include fractures other than hip, shoulder, clinical vertebral [spine], and forearm.

See Appendix D for full evidence tables.

1.1.5.4. GRADE profiles for bones assessment methods to predict fragility fracture

Table 3: GRADE profile for discriminatory power of bone assessment methods to predict major osteoporotic fracture

Bone assessment method Population, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	AUC (95% CI)[a]	GRADE certainty
DXA: aBMD-FN Mixed, N=75,734	12	Very serious[b]	Very serious[c]	Not serious	Not serious	0.65 (0.62-0.67)	VERY LOW
DXA: aBMD-FN – On treatment strata Postmenopausal women 55+ years (Briot 2013), N=1050	1	Serious[b]	Not serious	Not serious	Serious[d]	0.69 (0.61-0.76)	LOW
DXA: aBMD-LS Mixed, N=122,049	9[e]	Very serious[b]	Very serious[c]	Not serious	Not serious	0.63 (0.60-0.66)	VERY LOW
DXA: aBMD-TH Mixed, N=64,848	5	Very serious[b]	Very serious[c]	Not serious	Serious[d]	0.66 (0.61-0.71)	VERY LOW
DXA: aBMD-trochanter	1	Very serious[b]	Not serious	Serious[f]	Serious[d]	0.64 (0.57-0.72)	VERY LOW

Bone assessment method Population, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	AUC (95% CI)[a]	GRADE certainty
Postmenopausal women 65+ years (Popp 2009), N=637							
DXA: **aBMD-FN + TBS-LS** Men 50+ years referred for DXA (Leslie 2014), N=3630	1	Very serious[b]	Not serious	Not serious	Serious[e]	0.68 (0.64-0.72)	VERY LOW
DXA: **aBMD-LS + TBS-LS** Men 50+ years referred for DXA (Leslie 2014), N=3620	1	Very serious[b]	Not serious	Not serious	Not serious	0.64 (0.60-0.68)	LOW
QCT: **vBMD-LS** Adults hospitalised for fragility fracture (Leonhardt 2020), N=58	1	Very serious[b]	Not serious	Serious[f]	Serious[e]	0.76 (0.61-0.87)	VERY LOW
QUS-heel: BUA Women 65+ years	1	Very serious[b]	Not serious	Serious[g]	Serious[e]	0.71 (0.64-0.78)	VERY LOW

Bone assessment method Population, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	AUC (95% CI)[a]	GRADE certainty
(Chan 2013), N=312							
REMS: BMD-FN Women referred for DXA (Adami 2020), N=1370	1	Very serious[b]	Not serious	Serious[f]	Not serious	0.63 (0.58-0.67)	VERY LOW
REMS: BMD-LS Women referred for DXA (Adami 2020), N=1370	1	Very serious[b]	Not serious	Serious[f]	Not serious	0.63 (0.59-0.67)	VERY LOW

Abbreviations: aBMD, areal bone mineral density; AUC, area under the curve; BMD, bone mineral density; BUA, broadband ultrasonic attenuation; DXA, dual x-ray absorptiometry; FN, femoral neck; LS, lumbar spine; REMS, radiofrequency echographic multi spectrometry; TH, total hip; TBS, trabecular bone score; vBMD, volumetric bone mineral density.
Notes:

a. *when there are 3 or more studies, AUC point estimates and 95% CIs are from frequentist meta-analysis.*
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 QUADAS-2) or serious risk of bias in the evidence contributing to the outcomes (more than 50% of the weight of the evidence came from studies at moderate or high risk of bias as per QUADAS-2)*
c. *very serious inconsistency in the evidence contributing to the outcomes (visual inspection of forest plots, high I2, high τ2, and wide 95% prediction intervals) or serious inconsistency in the evidence contributing to the outcomes (high I2 but low τ2 due to small standard error of AUC but wide 95% prediction intervals).*
d. *very serious imprecision because 95% CI crosses 2 clinical decision thresholds (0.5 and 0.7) or serious imprecision because 95% CI crosses clinical decision threshold (0.5 or 0.7).*
e. *the 9 studies contribute 10 datasets as Schacter 2017B reported data for men and women separately.*
f. *study partially applicable as reference standard includes fragility fractures other than hip, shoulder, forearm, and clinical vertebral.*
g. *study partially applicable as reference standard only includes hip and clinical spine.*

Table 4: GRADE profile for discriminatory power of bone assessment methods to predict hip fracture

Bone assessment method Population, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	AUC (95% CI)[a]	GRADE certainty
DXA: aBMD-FN Mixed, N=94,434	11[b]	Very serious[c]	Very serious[d]	Not serious	Not serious	0.73 (0.70-0.76)	VERY LOW
DXA: aBMD-LS Mixed, N=120,725	6[e]	Very serious[c]	Not serious	Not serious	Not serious	0.66 (0.64-0.69)	LOW
DXA: aBMD-TH Mixed, N=75,738	5	Very serious[c]	Very serious[d]	Not serious	Not serious	0.76 (0.70-0.81)	VERY LOW
DXA: aBMD-total femur Postmenopausal women 75+ years (Hans 2004), N=5898	1	Very serious[c]	Not serious	Not serious	Not serious	0.64 (0.61-0.68)	LOW
DXA: aBMD-FN + TBS-LS Men 50+ referred for DXA (Leslie 2014), N=3620	1	Very serious[c]	Not serious	Not serious	Not serious	0.78 (0.71-0.85)	LOW
DXA: aBMD-LS + TBS-LS	1	Very serious[c]	Not serious	Not serious	Serious[f]	0.70 (0.62-0.78)	VERY LOW

Bone assessment method Population, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	AUC (95% CI)[a]	GRADE certainty
Men 50+ years referred for DXA (Leslie 2014), N=3620							
DXA: aBMD-trochanter Postmenopausal women 65+ years (Cummings 1994), N=7963	1	Very serious[c]	Not serious	Not serious	Not serious	0.78 (0.73-0.83)	LOW
QUS-heel: BUA[g] Postmenopausal women 65+ (Chan 2013), N=312	1	Very serious[c]	Not serious	Not serious	Not serious	0.85 (0.75-0.95)	LOW
QUS-heel: BUA[h] Postmenopausal women 75+ (Hans 2004), N=5898	1	Very serious[c]	Not serious	Not serious	Not serious	0.65 (0.62-0.69)	LOW
QUS-heel: BUA[i,j] Postmenopausal women, 70+ (Krieg 2006), N=7062	1	Serious[c]	Not serious	Not serious	Serious[f]	0.71 (0.66-0.77) 0.72 (0.67-0.78)	LOW

Bone assessment method Population, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	AUC (95% CI)[a]	GRADE certainty
QUS-heel: SOS[h] Postmenopausal women 75+ (Hans 2004), N=5898	1	Very serious[c]	Not serious	Not serious	Not serious	0.65 (0.61-0.68)	LOW
QUS-heel: SOS[i,j] Postmenopausal women, 70+ (Krieg 2006), N=7062	1[k]	Serious[c]	Not serious	Not serious	Serious[f]	0.70 (0.64-0.76) 0.72 (0.66-0.78)	LOW
QUS-heel: composite (BUA and SOS)[h] Stiffness index Postmenopausal women, 70+ (Hans 2004), N=12,960	1	Very serious[c]	Not serious	Not serious	Serious[f]	0.66 (0.62-0.70)	VERY LOW
QUS-heel: composite (BUA and SOS)[j] Quantitative Ultrasound Index Postmenopausal women, 70+	1[k]	Serious[c]	Not serious	Not serious	Serious[f]	0.72 (0.66-0.78) 0.73 (0,67-0.79)	LOW

Bone assessment method Population, number of participants	Number of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	AUC (95% CI)[a]	GRADE certainty
(Krieg 2006), N=7062							
REMS: BMD-FN Women referred for DXA (Adami 2020), N=1370	1	Very serious[c]	Not serious	Not serious	Very serious[f]	0.60 (0.48-0.72)	VERY LOW
REMS: BMD-LS Women referred for DXA (Adami 2020), N=1370	1	Very serious[c]	Not serious	Not serious	Serious[f]	0.66 (0.54-0.77)	VERY LOW

Abbreviations: aBMD, areal bone mineral density; BMD, bone mineral density; BUA, broadband ultrasound attenuation; DXA, dual x-ray absorptiometry; FN, femoral neck; LS, lumbar spine; QUI, quantitative ultrasound index; SI, stiffness index; SOS, speed of sound; TH, total hip.

Notes:

a. *When there are 3 or more studies, AUC point estimates and 95% CIs are from frequentist meta-analysis.*
b. *The 11 studies contribute 12 datasets as Ensrud 2024 reports data for men and women separately.*
c. *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 QUADAS-2) or serious risk of bias in the evidence contributing to the outcomes (more than 50% of the weight of the evidence came from studies at moderate or high risk of bias as per QUADAS-2).*
d. *Very serious inconsistency in the evidence contributing to the outcomes (visual inspection of forest plots, high I^2, high τ^2, and wide 95% prediction intervals) or serious inconsistency in the evidence contributing to the outcomes (high I^2 but low τ^2 due to small standard error of AUC but wide 95% prediction intervals).*
e. *The 6 studies contribute 7 datasets as Schacter 2017 reported data for men and women separately.*
f. *Very serious imprecision because 95% CI crosses 2 clinical decision thresholds (0.5 and 0.7) or serious imprecision because 95% CI crosses clinical decision threshold (0.5 or 0.7).*
g. *QUS measure obtained using McCue CUBAClinical sonometer.*
h. *QUS measure obtained with GE Lunar Achilles sonometer.*
i. *QUS measure obtained with GE Lunar Achilles+ sonometer.*
j. *QUS measure obtained with Hologic Sahara sonometer.*
k. *Study contributes 2 datasets as Krieg 2006 reported data for 2 sonometers.*

1.1.5.5. Sensitivity and specificity of bone assessment methods

Table 5: Sensitivity and specificity of DXA measurement of BMD at the femoral neck at specific fracture risk thresholds to predict hip fracture

Study	Number of participants	Threshold	Sex	Sensitivity	Specificity
Center 2004	926	T-score ≤ -2.5 SD	Women	0.73 (0.61-0.82)	0.81 (0.78-0.84)
	658	T-score ≤ -2.5 SD	Men	0.43 (0.23-0.66)	0.92 (0.90-0.94)
Dargent-Molina 2003	5910	BMD threshold below which fracture risk>20 per 1000 woman-years	Women	0.35 (0.29-0.41)	0.86 (0.85-0.87)
Dargent-Molina 1999 [a]	5895	Top 50% highest risk (≥50% fracture risk)	Women	0.85 (0.79-0.90)	NR
	5895	Top 25% highest risk (≥75% fracture risk)	Women	0.52 (0.44-0.59)	NR
	5895	Top 10% highest risk (≥90% fracture risk)	Women	0.29 (0.22-0.36)	NR

Notes:a, Dargent-Molina 1999 did not report specificity and reports results adjusted for age.

Table 6: Sensitivity and specificity of QUS measurement of BUA at the heel (calcaneus) at specific fracture risk thresholds to predict hip fracture

Study	Number of participants	Threshold	Sex	Sensitivity	Specificity
Dargent-Molina 2003	5910	BMD threshold below which fracture risk>20	Women	0.15 (0.11-0.20)	0.95 (0.94-0.96)

		per 1000 woman-years			
Dargent-Molina 1999[a]	5895	Top 50% highest risk (≥50% fracture risk)	Women	0.74 (0.67-0.81)	NR
	5895	Top 25% highest risk (≥75% fracture risk)	Women	0.55 (0.47-0.63)	NR
	5895	Top 10% highest risk (≥90% fracture risk)	Women	0.29 (0.22-0.36)	NR

Notes:a, Dargent-Molina 1999 did not report specificity and results are adjusted for age.

1.1.6. Economic evidence

Economic evidence related to bone assessment methods was sought as part of evidence review E. No included studies compared alternative bone assessment methods.

1.1.7. Economic model

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

1.1.8. Unit costs

The Unit cost for a bone mineral density (BMD) assessment using DXA is reported below.

Table 7: Costs associated with BMD assessment with DXA scan

Resource	Cost	Source
BMD assessment using DXA scan	£84[(a)]	NHS National Cost Collection 2023/24

Abbreviations: BMD= bone mineral density, DXA= dual-energy X-ray absorptiometry
(a) Weighted average cost of DXA (Currency code RD40Z).

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 bone assessment methods should predict were major osteoporotic (MOF) fracture and hip fracture (HF). MOF was defined as hip, clinical vertebral (spine), humerus and forearm in accordance with the FRAX and QFracture (2012) risk prediction tools. Some studies reported fracture outcomes that included fractures not within the definition of MOF.

The following statistical outcomes were identified as relevant to assessing the performance of bone assessment methods:

- Area under the curve (AUC) for overall discriminatory power.
- Sensitivity, specificity, and predictive values for discrimination at specific thresholds.

Most of the included studies reported AUC while only three studies reported sensitivity and specificity outcomes for discrimination of HF at specific risk thresholds.

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

Evidence was identified for DXA, CT and ultrasound that assessed the accuracy of these imaging methods to predict MOF and HF. No evidence was identified that assessed the accuracy of digital radiography to predict fragility fracture.

Only the most recently published article of the Australian DOES cohort (Center 2004) was included to prevent double counting of cohort participants because some earlier articles covered the same bone assessment method and measure at an overlapping time period.

Most of the studies included in the prediction of MOF used the definitions specified above, although some were more permissive and included other types of fragility fracture. In these cases, the studies were included but downgraded for indirectness.

The accuracy results for the Bayesian model were similar to those of the frequentist model. In general, the AUC point estimates and 95% CIs, as estimated by a frequentist model, for DXA measurement of aBMD at the various skeletal sites (femoral neck, lumbar spine, total hip) for predicting MOF and HF were like those produced by a Bayesian model. Results from the frequentist model were presented because those from the Bayesian meta-analysis models were similar, although the 95% CIs and prediction intervals were slightly wider, reflecting the uncertainty in the evidence.

In line with the protocol, heterogeneity was explored according to treatment history at baseline and follow up time. Follow-up time ranged from 2 years to >10 years. Initial investigation of follow up time as a covariate by meta regression did not suggest a strong correlation with the AUC estimates and so was not further explored. For treatment history, heterogeneity was examined when there were two or more studies in each group contributing such data: measurement of aBMD at the femoral neck to predict 10-year MOF and HF. Similar investigation of treatment history at baseline as a covariate by meta regression did not suggest a strong correlation with the AUC estimates and was also not further explored.

For the outcome of MOF, most of the bone assessment methods were assessed as very low certainty. All the included studies were at overall moderate or high risk of bias. Six studies (Adami 2020, Chan 2013, Chapurlat 2020, Crandall 2020, Kuzma 2018, Leslie 2014) were at high risk of bias due to concerns about patient selection, such as unclear reporting of recruitment methods or inappropriate exclusions from the samples (for example, people with osteoporosis). Most studies were downgraded for the flow and timing of patient assessment. This included when radiographs of patients were only checked to confirm fracture when this was self-reported. Most studies were downgraded for reference standard bias as participants did not all receive the same reference standard (for example, because only the radiographs of participants who self-reported fracture were checked) were downgraded due to the potential for verification bias. Five studies (Adami 2020, Chan 2013, Kuzma 2018, Leonhardt 2020, Popp 2009) were downgraded for indirectness of the reference standard because they included more types of fracture than typically included in the category of ‘major osteoporotic fracture’ and these may therefore overestimate accuracy of a bone assessment method.

There was very serious inconsistency in the estimated AUC of the bone assessment methods for which there was sufficient evidence to conduct meta-analysis: DXA of the femoral neck, lumbar spine, and total hip. The forest plots showed wide variation in results, and there was high I^2 (>75%) and τ^2 (>0.01) values, and wide 95% prediction intervals. This indicates that there is still substantial uncertainty in what the discriminatory power of DXA of the femoral neck, lumbar spine and total hip will be in future studies and contexts. Outcomes were downgraded for imprecision when confidence intervals crossed one of the clinical decision thresholds. The wide 95% prediction intervals in all cases were consistent with the other measures suggesting substantial heterogeneity in the results.

For the outcome of HF, most of the identified bone assessment methods were low certainty for similar reasons as for MOF, although many of them were single studies. Unlike for MOF, studies that ascertained fractures by self-report only were assessed as having a low risk of indirectness as it is unlikely that these fractures would be missed. All meta-analysed bone assessment methods for prediction of HF, except for that of DXA measurement of aBMD at the lumbar spine, had very serious inconsistency in the estimated AUC all of which exhibited wide variation of results on the forest plots, high I^2 (>75%) and τ^2, and wide 95% prediction intervals. The wide 95% prediction intervals in all cases were consistent with the other measures suggesting substantial heterogeneity in the results.

In general, there were various other potential unexplored sources of heterogeneity in the evidence, including: whether reported BMD values were adjusted for age, unadjusted, or adjusted for other factors; and whether the population was at risk, suspected risk, had previous fracture, or were unselected post-menopausal women.

1.2.3. Benefits and harms

1.2.3.1. DXA

Overall accuracy for prediction of MOF

The AUC estimates for the overall accuracy of aBMD as measured by DXA to predict MOF suggest that they have poor discriminatory power across all possible thresholds in adults at

risk or suspected risk of fragility fracture. The results for the 3 meta-analyses of aBMD measured by DXA at the femoral neck, lumbar spine (L1-L4), and total hip were all within the 0.6-0.7 range. The skeletal site with the highest point estimate obtained from meta-analysis was total hip (0.66 [95%CI 0.61-0.71]). This means that, given two randomly picked patients, one with and one without MOF, DXA of aBMD at the total hip will assign a higher probability of developing a MOF to the person with a fracture 66% of the time. One study of a small cohort (Briot 2013) in a subgroup of patients receiving anti-osteoporosis treatment found that DXA measurement of aBMD at the femoral neck had a similar accuracy of 0.69 (95%CI 0.61-0.76) although there was similar uncertainty. Single studies reported DXA at the trochanter (Popp 2009) or combined with a trabecular bone score assessment at the lumbar spine and femoral neck (Leslie 2014) found similar poor discriminatory power ranging from 0.64-0.68.

Given the poor performance overall of aBMD as measured by DXA to predict MOF, it is likely that a substantial number of people would be incorrectly classified as either at high risk or not at risk of sustaining a major osteoporotic fracture.

Overall accuracy for prediction of hip fracture

As for MOF, the estimates for the overall accuracy of DXA measurement of aBMD to predict HF suggest that they potentially have poor to moderate discriminatory power across all possible thresholds in adults at risk or suspected risk of HF indicating that it is likely that a substantial number of people would be incorrectly classified as either developing or not developing a hip fracture.

The AUC results for the 3 meta-analyses of DXA measurement of femoral neck and total hip ranging from 0.73-0.76 and a point estimate for lumbar spine (L1-L4) of 0.66. The skeletal site with the highest point estimate obtained from meta-analysis was total hip (0.76 [95%CI 0.70-0.81]). This means that, given two randomly picked patients, one with and one without HF, DXA of the total hip will assign a higher probability of developing a HF to the person with a fracture 76% of the time.

Single studies reported DXA at the trochanter (Cummings 1994), total femur (Hans 2004) or combined with a trabecular bone score assessment at the lumbar spine and femoral neck (Leslie 2014) found poor to moderate discriminatory power ranging from 0.64-0.78.

Sensitivity and specificity at specific fracture risk threshold for hip fracture

Sensitivity and specificity of DXA measurement of aBMD at the femoral neck using various risk thresholds to predict hip fracture was reported in 3 studies (Center 2004, Dargent-Molina 1999, Dargent-Molina 2004). The results showed that generally sensitivity appeared to decrease as the fracture risk threshold was increased (leading to more false negatives and therefore missed fracture patients). Only two of the three studies (Center 2004, Dargent-Molina 2003) explicitly reported specificity at the chosen thresholds, which appeared to be generally good (>0.8). One study (Dargent-Molina 2003) reported sensitivity and specificity for men and women separately. The women had considerably higher sensitivity then the men (73% versus 43% respectively). The third study (Dargent-Molina 1999) did not reported specificity on the grounds that it appeared to be approximately the same as the fracture risk threshold used (for example, specificity of 0.75 for fracture risk threshold of 75%). The sensitivity reported was adjusted for age.

1.2.3.2. QCT scan

Overall accuracy to predict MOF

Measurement of volumetric BMD (vBMD) by quantitative computed tomography (QCT) of the lumbar spine to predict major osteoporotic fracture was examined in one study (Leonhardt 2020) and had an AUC of 0.76 (95%CI 0.61-0.87), the highest of all the bone assessment methods examined in this review. However, this was a very small study (n=58) conducted in

adults hospitalised for fragility fracture and the wide confidence intervals indicate there is substantial uncertainty.

1.2.3.3. Quantitative ultrasound

Overall accuracy to predict MOF

Two studies assessed the overall discriminatory power of QUS at the heel (that is, the calcaneus) to predict major osteoporotic fracture. One small study (N=312; Chan 2013) found that the accuracy of BUA measured by QUS at the heel to predict MOF in postmenopausal women ≥ 65 years was moderate with an AUC of 0.71 although there was substantive uncertainty (95% CI 0.64-0.78).

Measurement of aBMD using radiofrequency echographic multi spectrometry (REMS), a more recent form of QUS, was estimated by one study in white women aged 30-90 years referred for a DXA (N=1516; Adami 2020) to have poor discriminatory power of 0.63 at both the femoral neck (95% CI 0.58-0.67) and the lumbar spine (95% CI 0.59-0.67).

1.2.3.4. Overall accuracy to predict HF

Three studies assessed the overall discriminatory power of QUS at the heel to predict hip fracture (Chan 2013, Hans 2004, Kreig 2006). The discriminatory power of the various QUS measures ranged from poor to good (AUC of 0.65 to 0.85) for BUA and poor to moderate (0.65 to 0.72) for SOS with the related 95% CIs indicating some uncertainty. The composite measure combining BUA and SOS also ranged from 0.66 to 0.73 and did not appear to substantially improve the accuracy of the single measures.

Measurement of aBMD using radiofrequency echographic multi spectrometry (REMS), a more recent form of QUS, was estimated by one study (N=1516; Adami 2020) in white women aged 30-90 years referred for a DXA scan to have poor discriminatory power at the femoral neck of 0.60 (95% CI 0.48-0.67) and poor discriminatory power at the lumbar spine of 0.66 (95% CI 0.54-0.77) in white women aged 30-90 years referred for a DXA scan. However, the wide confidence intervals indicate substantive uncertainty.

Sensitivity and specificity at specific fracture risk threshold to predict hip fracture

Sensitivity and specificity of DXA measurement of aBMD at the femoral neck using various risk thresholds to predict hip fracture was reported in 2 studies (Dargent-Molina 1999, Dargent-Molina 2004). The results were like those for the prediction of MOF with sensitivity appearing to decrease as the fracture risk increases (leading to more false negatives and missed fracture patients). Specificity was reported in one study (Dargent-Molina 2003) and was excellent at 0.95; the other study (Dargent-Molina 1999) did not report specificity for the same reasons discussed above and adjusted results for age.

1.2.3.5. Comparison of MOF and HF

Overall, the discriminatory power of bone assessment methods to predict MOF and HF were similar for DXA, CT and ultrasound. The AUC point estimates for DXA of the various skeletal sites were higher for HF than for MOF and suggest that they have at best good but potentially only moderate discriminatory power across all possible thresholds in adults at risk or suspected risk of fragility fracture. For MOF and HF, the skeletal site with the highest point estimate obtained from meta-analysis was total hip.

1.2.4. Conclusions and committee experience

The committee recognised that measurement of BMD is an important piece of information when assessing fracture risk, and making decisions about whether treatment is clinically appropriate, and monitoring risk and treatment progress.

The committee acknowledged that most identified evidence was on the discriminatory power of aBMD as measured by DXA at the femoral neck or lumbar spine, which are the most imaged skeletal sites, to predict fracture, with very few studies identified at other skeletal sites and on QCT or QUS. The poor to moderate performance of DXA measurement of aBMD means that the potential to miss fractures and to wrongly predict fracture is high. Therefore, the committee agreed that measurement of an individual's aBMD using DXA should not be used on its own to predict fracture risk but rather should be interpreted in the context of their fracture risk profile, including: their fracture history, estimated fracture risk (as estimated by FRAX or QFracture; see Evidence review C) and consideration of the individual's risk factors (see Evidence review A), to decide treatment options.

The evidence for the other bone assessment methods to predict fragility fracture was limited and the committee agreed that no recommendations should be made about their use to predict who will develop fragility fracture and who will not.

Measurement of BMD by DXA rather than other bone assessment techniques was recommended from this review because it had the most evidence and it is widely used in current practice. Further recommendations about when to consider a DXA are discussed in Evidence review E.

1.2.4.1. Research recommendation

There was only one identified study on REMS, which is a more recent ultrasound-based and non-ionising radiation imaging technology, to predict fracture. The committee agreed that benefits of this technology compared to DXA are that it is quick, safe, and does not require special arrangements. Since REMS does not involve exposure to ionising radiation, its use does not require shielding to protect the patient. Furthermore, REMS can provide additional data about bone health/quality not restricted to aBMD and the machines (unlike DXA machines) are easily portable allowing imaging in a variety of settings. The committee agreed that further studies are needed to assess the ability of REMS to predict fragility fracture in adults at risk or suspected risk of fragility fracture. A research recommendation was therefore proposed to explore the discriminatory power of this bone assessment method.

1.2.5. Cost effectiveness and resource use

Choice of bone assessment technique

No cost-effectiveness analyses were included in review E that compared different bone assessment techniques.

Current practice is to use DXA for bone assessment and the committee did not consider there to be sufficient clinical evidence to support use of alternatives and so costs of alternatives were not considered.

The committee noted that some areas have capacity issues with DXA and so there can be long waiting lists that delay treatment. They agreed that DXA capacity should be sufficient for timely assessment. NHS England data from November 2024 reported a median waiting time for DXA of around 2.8 weeks but that 19% of people on the waiting list had been so for 6 weeks or more and 6% for 13 weeks or more. They noted that NHS England has already committed to funding 13 DXA scanners to increase capacity.

1.2.6. Other factors the committee took into account

The committee discussed the lack of evidence for women with pregnancy and lactation-associated osteoporosis but acknowledged that this was outside the scope of this guideline.

1.2.7. Recommendations supported by this evidence review

These evidence reviews support recommendations 1.4.1-1.4.4 and the recommendation for research on the accuracy of radiofrequency echo-graphic multi spectrometry (REMS) for predicting fragility fractures in adults. There is overlap between evidence reviews and recommendations from evidence reports C, D and E.

1

2 # 1.3. References

1.3.1. Clinical

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

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

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: the official journal of the American Society for Bone and Mineral Research 26(2): 420-7

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1.3.2. Other

Debray, Thomas PA; Damen, Johanna AAG; Riley, Richard D 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.

Debray, Thomas PA; Damen, Johanna AAG; Snell, Kym IE et al. (2017) A guide to systematic review and meta-analysis of prediction model performance. BMJ 356: i6460.

Appendices

Appendix A Review protocols

A.1 Review protocol for What is the accuracy of bone assessment methods for predicting fragility fractures in adults including those who have had a previous fragility fracture?

Field	Content
Review title	Bone assessment methods to predict fragility fracture
Review question	What is the accuracy of bone assessment methods for predicting 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 Amendment to protocol: people on drug treatment was not intended as a population stratum but as a sub-group for heterogeneity. It has been removed from the protocol post analysis.

Risk prediction tool/bone assessment method	The following methods to assess bone density and quality to predict MOF and HF will be included:
	• Dual X-ray absorptiometry (DXA, DEXA) or dual x-ray and laser (DXL) of hip, spine, or forearm
	  ○ Areal BMD (aBMD) only
	  ○ aBMD with TBS assessment
	• Quantitative computed tomography scans (QCT), including asynchronous calibration QCT (phantom-less scanning); high-resolution peripheral QCT (HR-pQCT); peripheral QCT (pQCT); and photon-counting CT
	  ○ Volumetric BMD (vBMD)

	• Quantitative ultrasound (QUS) (for example Bindex)
	  ○ Broadband ultrasound attenuation (BUA)
	  ○ Speed of sound (SOS)
	  ○ Quantitative ultrasound index (QUI) or Stiffness Index (SI) (both combine BUA and SOS measurement)
	• Digital radiography (IBEX BH software)
	  ○ aBMD
	Bone assessment methods do not require validation in a UK-only population as there is little variation in bone mineral density between countries. Gold/reference standard is combination of clinical review, self-report, and confirmation of fracture by radiography. QUS measurements can vary substantially between machines, therefore results will be presented by type of machine.
	Note: This is an amendment to the initial protocol, undertaken after the initiation of data analysis, to clarify the following bone assessment methods:
	- Include details about the reference standard used to assess the bone assessment methods

Target condition	Fragility fracture. • Major osteoporotic fracture (MOF) • Hip fracture (HF)

Types of study to be included	**Inclusion:** • Prospective cohort studies **Exclusion:** • Retrospective cohort studies • Case-control studies • Cross-sectional studies • Studies using machine learning algorithms, polygenic risk scores, or radiomic models will be excluded
Other exclusion criteria	• Non-English language studies • Conference abstracts
Context	All settings
Primary outcomes (critical outcomes)	The following measures of test accuracy will be used: Overall discrimination between fragility fracture and no fragility fracture • c-statistic/AUC Discrimination between fracture and no fragility fracture at a specific threshold • 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: • QUADAS-2 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). Note: This is an amendment to the initial protocol, undertaken after the initiation of data analysis, to clarify that QUADAS-2 will be used to assess risk of bias.
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		
Type and method of review	☐	Intervention	
	☐	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] 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.

Q3.1b What is the accuracy of bone assessment methods for predicting fragility fractures in adults including those who have had a previous fragility fracture?

Table 8: 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

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

1

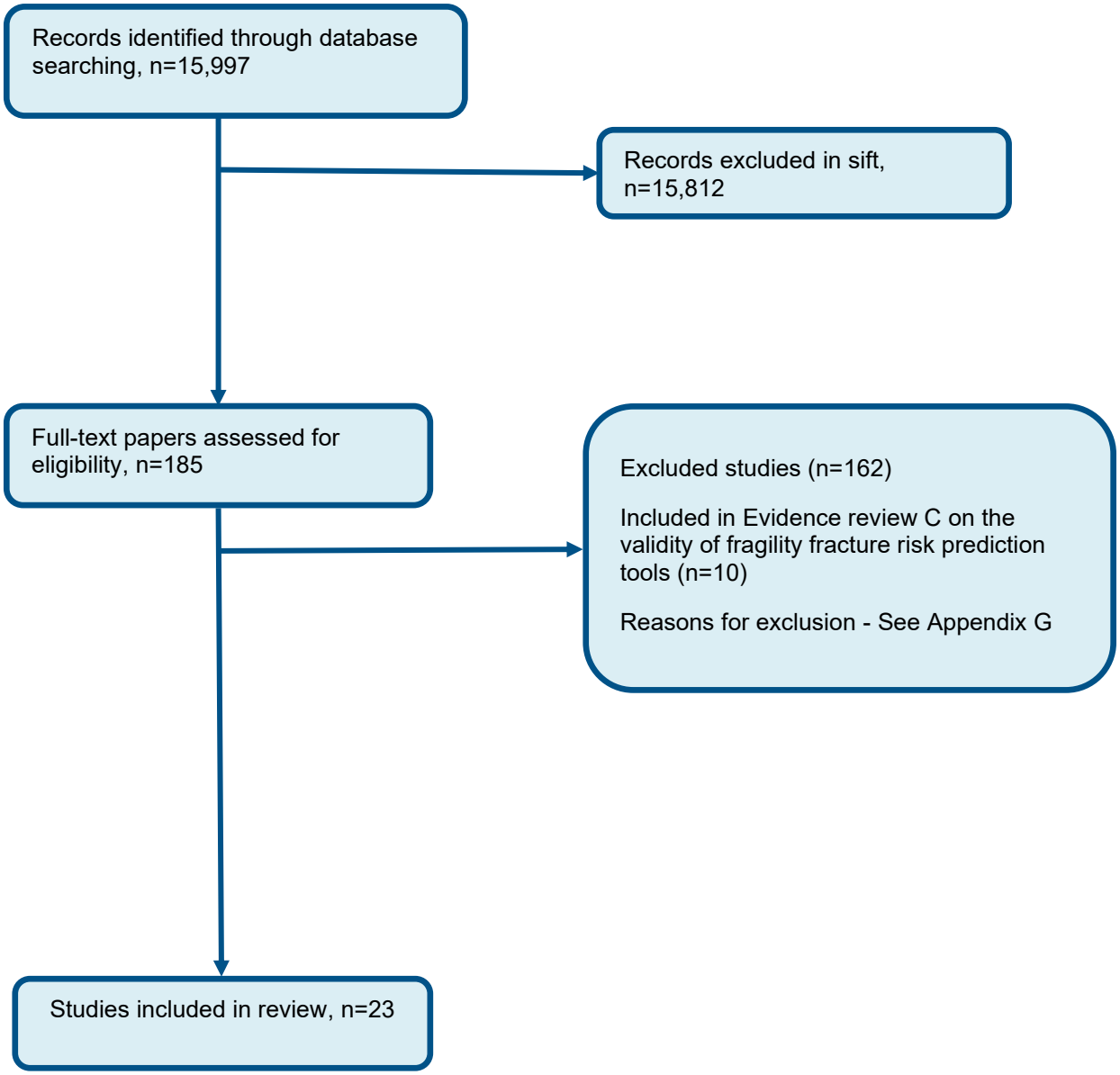
Epistemonikos search terms

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 Bone assessment methods evidence study selection

Figure 1: Flow chart of clinical study selection for the risk prediction and bone assessment methods reviews



Appendix D Bone assessment methods evidence

D.1.1 Adami, 2020

Bibliographic Reference Adami, Giovanni; Arioli, Giovanni; Bianchi, Gerolamo; Brandi, Maria Luisa; Caffarelli, Carla; Cianferotti, Luisella; Gatti, Davide; Girasole, Giuseppe; Gonnelli, Stefano; Manfredini, Monica; Muratore, Maurizio; Quarta, Eugenio; Quarta, Laura; Radiofrequency echographic multi spectrometry for the prediction of incident fragility fractures: A 5-year follow-up study.; Bone; 2020; vol. 134; 115297

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	Prospective cohort study
Study location	Italy
Study setting	Secondary (Galateo Hospital in San Cesario di Lecce, Italy)

Study dates	Recruited between October 2013 and October 2016
Sources of funding	Not reported
Study sample	N=1516 Caucasian women aged 30–90 years On treatment at baseline: NR Received treatment during FU: NR
Inclusion criteria	• Caucasian ethnicity • female sex • aged 30–90 years • absence of significant deambulation impairment • medical prescription for an axial DXA investigation • provision of written informed consent
Exclusion criteria	Not reported
Bone assessment method	Quantitative ultrasound Radiofrequency echographic multi spectrometry BMD-FN T-score; BMD-LS T-score Dual X-ray absorptiometry BMD-FN T-score; BMD-LS T-score
Bone assessment method - baseline	Participants had DXA scans performed using a Discovery W (Hologic, Waltham, MA, USA) scanner according to the standard clinical routine procedures. Spinal investigations were carried out with hip and knee both at 90° of flexion, and during femoral scans the patient's femur was straight on the table, with the shaft being parallel to the vertical edge of the obtained image, with a 15–25° internal rotation. The DXA scanner underwent daily quality control and regular maintenance for the whole study period

Bone assessment method - length of follow up	5-year follow-up period, with Mean=3.7 years (SD 0.8)
Bone assessment method - follow up	Reference standard: Medical reports based on imaging (for example, radiographs or vertebral morphometry) Follow-up was completed in 1,370 participants (146 patients (9.6%) voluntarily dropped out from the study or died). Every 6 months, participants were contacted by telephonic interview to assess their health status. Declared fractures were verified by medical reports based on imaging investigations, such as radiographs and vertebral morphometry. Participants that suffered more than one fragility fracture were not excluded from the study but only the first occurred fracture was considered in the analysis. Traumatic fractures were excluded.
Predicted outcomes	• Major osteoporotic fracture • Hip fracture *Note: data for MOF includes wrist, vertebra, shoulder, hip, ribs, forearm, ankle, pelvis, and other fractures.*
Discrimination outcomes	c-statistic/AUC

Study arms

- Major osteoporotic fracture (N = 192)
- No major osteoporotic fracture (N = 1178)
- Hip fracture (N = 21)
- No hip fracture (N = 1349)

Characteristics

Study-level characteristics

Characteristic	Study (N = 1370)
% Female No of events	n = 1370; % = 100
Mean age (SD) Range	54 to 60
Mean age (SD) Mean (SD)	60 (NR)
BMI (kg/m^2) Range	22.3 to 26.6
BMI (kg/m^2) Mean (SD)	24 (NR)

Outcomes

DXA BMD-FN t-score

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 1178, N1 = 192	Hip fracture vs No hip fracture, N2 = 1349, N1 = 21
AUC Adjusted for age. Hip fracture also adjusted for BMI. Custom value	0.583 (95%CI 0.539-0.626)	0.616 (95%CI 0.491-0.727)

1 AUC - Polarity - Higher values are better

2 AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group.
3 95%CIs calculated using equations in Section 3.1, Debray 2018.

4 **DXA BMD-LS t-score**

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 1178, N1 = 192	Hip fracture vs No hip fracture, N2 = 1349, N1 = 21
AUC Age adjusted. Hip fracture also adjusted for BMI. Custom value	0.597 (95%CI 0.553-0.639)	0.674 (95%CI 0.550-0.777)

5 AUC - Polarity - Higher values are better

6 L1-L4.

7 AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group.
8 95%CIs calculated using equations in Section 3.1, Debray 2018.

9 **REMS BMD-FN t-score**

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 1178, N1 = 192	Hip fracture vs No hip fracture, N2 = 1349, N1 = 21
AUC Adjusted for age. Hip fracture also adjusted for BMI. Custom value	0.627 (95%CI 0.584-0.668)	0.602 (95%CI 0.477-0.715)

10 AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group.
11 95%CIs calculated using equations in Section 3.1, Debray 2018.

1 **REMS BMD-LS t-score**

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 1178, N1 = 192	Hip fracture vs No hip fracture, N2 = 1349, N1 = 21
AUC Age adjusted. Hip fracture also adjusted for BMI. Custom value	0.631 (95%CI 0.588-0.672)	0.664 (95% CI 0.54-0.769)

2 AUC - Polarity - Higher values are better

3 AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group.
4 95%CIs calculated using equations in Section 3.1, Debray 2018.

5 **Critical appraisal - Critical Appraisal - QUADAS-2**

Overall risk of bias and directness	Risk of Bias	High *(Study assessed as high RoB for hip fracture due to high risk of bias in the domains patient selection and flow and timing. This is because it is unclear whether a random sample or consecutive participants were enrolled, and the exclusion criteria was not reported. Not clear whether all patients received same reference standard due to insufficient information reported. Study assessed as high RoB for MOF due to additional concerns about the reference standard as insufficient information about fracture ascertainment reported.)*
Overall risk of bias and directness	Directness	Partially applicable for MOF *(Includes all fragility fractures (wrist, spine, shoulder, hip, ribs, forearm, ankle, pelvis, and other types). Fracture of ribs, forearm, ankle, pelvis, and other types comprise ~50% of all fractures in sample).* Directly applicable for hip fracture.

6 ### D.1.2 Azagra, 2016

Bibliographic Reference Azagra, Rafael; Zwart, Marta; Encabo, Gloria; Aguye, Amada; Martin-Sanchez, Juan Carlos; Puchol-Ruiz, Nuria; Gabriel-Escoda, Paula; Ortiz-Alinque, Sergio; Gene, Emilio; Iglesias, Milagros; Morina, David; Diaz-Herrera, Miguel Angel; Utzet,

Mireia; Manresa, Josep Maria; Rationale of the Spanish FRAX model in decision-making for predicting osteoporotic fractures: an update of FRIDEX cohort of Spanish women.; BMC musculoskeletal disorders; 2016; vol. 17; 262

1 **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	Prospective cohort study
Study location	Spain
Study setting	Community
Study dates	2000 to 2010
Sources of funding	Supported in part by research grants from FEDER (European Union), Instituto de Salud Carlos III, Ministry of Economy and Competitivity and the Institut Universitari d'Investigació en Atenció Primària IDIAP Jordi Gol, and Cátedra UAB-Novartis 2009 Scholarship.
Study sample	N=1918 women were contacted and N=1308 women aged ≥40 and ≤90 years were included in the analysis On treatment at baseline: No Received treatment during FU: NR
Inclusion criteria	• Caucasian women aged ≥40 and ≤90 years

	• understood and spoke Spanish • responded to both the initial questionnaire on risk factors and 10-year follow-up telephone questionnaire • enrolled in FRIDEX cohort
Exclusion criteria	• refused informed consent to participate in the study • without a telephone contact number • did not respond after 3 attempted phone calls • physical or psychological difficulties that prevented participation in the study, or relatives who refused permission for the individual to participate • Paget's disease or bone cancer
Bone assessment method	Dual X-ray absorptiometry BMD-FN T-score
Bone assessment method - baseline	Participants underwent axial bone densitometry DXA and a questionnaire on risk factors for osteoporotic fracture at baseline.
Bone assessment method - length of follow up	≥10 years
Bone assessment method - follow up	Reference standard: Self-report with fracture confirmed by medical records Follow-up was completed in 1,308 women (A total of 1,918 women were contacted at the end of the 10-year period and 86 subjects refused to participate, 33 were excluded due to cancer and 491 were excluded because they had been receiving anti-osteoporotic medication at baseline). Fractures that could not be confirmed were excluded from the analysis. The major osteoporotic fractures (hip, humerus, forearm, and clinical vertebral [spine]) during the follow up period were taken as the endpoint event.

Predicted outcomes	• Major osteoporotic fracture (Hip, shoulder, clinical vertebral [spine], forearm fractures) • Hip fracture
Discrimination outcomes	c-statistic/AUC

Study arms

- Major osteoporotic fracture (N = 108)
- No major osteoporotic fracture (N = 1200)
- Hip fracture (N =26)
- No hip fracture (N = 1282)

Characteristics

Study-level characteristics

Characteristic	**Study (N = 1918)**
% Female No of events	n = 1918 ; % = 100
Mean age (SD) Mean (SD)	57.5 (8.1)
Ethnicity	n = NA ; % = NA

Characteristic	Study (N = 1918)
No of events	
Ethnicity - Caucasian No of events	n = 1918 ; % = 100
BMI (kg/m^2) Mean (SD)	27.7 (4.6)
Alcohol intake No of events	n = NA ; % = NA
Alcohol intake - High alcohol intake No of events	n = 134 ; % = 0.7
Smoking status - Current smoking No of events	n = 207 ; % = 10.8
Previous fracture - Personal history of fractures No of events	n = 433 ; % = 22.6
Rheumatoid arthritis or systemic lupus erythematosus No of events	n = NA ; % = NA
Rheumatoid arthritis or systemic lupus erythematosus - Rheumatoid arthritis	n = 25 ; % = 1.3

Characteristic	Study (N = 1918)
No of events	
Corticosteroid use - Glucocorticoids No of events	n = 90 ; % = 4.7

Data is for randomly selected patients who had completed at least 10-year FU and who responded to FU

Outcomes

DXA BMD-FN t-score

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 1200, N1 = 108	Hip fracture vs No hip fracture, N2 = 1282, N1 = 26
AUC Unadjusted Custom value	0.706 (95%CI 0.652-0.76)	0.814 (95%CI 0.712-0.916)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	Outcome MOF

		High *(Study at high RoB for MOF due to concerns about reference standard and flow and timing. Although self-reported incident fragility fractures confirmed by medical records, no information about whether radiographs/radiographic reports consulted. Not all participants received same reference standard (medical records checked only if patient reported fracture)* *Outcome HF* Moderate *(Study at moderate RoB for HF due to concerns about flow and timing of test as described above.)*
Overall risk of bias and directness	Directness	Directly applicable

D.1.3 Bolland, 2011

Bibliographic Reference Bolland, Mark J; Siu, Amanda Ty; Mason, Barbara H; Horne, Anne M; Ames, Ruth W; Grey, Andrew B; Gamble, Greg D; Reid, Ian R; Evaluation of the FRAX and Garvan fracture risk calculators in older women.; Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research; 2011; vol. 26 (no. 2); 420-7

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	Prospective cohort study
Study location	New Zealand
Study setting	Not reported
Study dates	The study started in 1998 and was completed in 2005. Between mid-2008 and mid-2009 fractures and other medical events since study completion were recorded
Sources of funding	Grants from the Health Research Council of New Zealand
Study sample	N=1422 healthy postmenopausal women On treatment at baseline: No Received treatment during FU: NR
Inclusion criteria	• women older than 55 years of age • free from major medical conditions • normal lumbar spine bone mineral density (BMD) for their age (Z-score > –2) • not taking treatment for osteoporosis (including hormone replacement therapy or vitamin D supplements) in doses > 1000 IU/day • had serum 25(OH)D levels ≥ 25 nmol/L
Exclusion criteria	• No measurement of femoral neck BMD at baseline
Population subgroups	

Bone assessment method	Dual X-ray absorptiometry BMD-FN T-score
Bone assessment method - baseline	Medical history was obtained by questionnaire, weight was measured using electronic scales, height was measured using a Harpenden stadiometer, and BMD of the femoral neck was determined using a Lunar Expert dual-energy X-ray absorptiometer, software Version 1.7 (GE Lunar, Madison, WI, USA)
Bone assessment method - length of follow up	Mean=8.8 years (range 0.2 to 11.4, total follow-up 12,500 patient-years)
Bone assessment method - follow up	Reference standard: Self-report with fracture confirmed by radiographs/radiographic review (original 5-year RCT). Self-report (subsequent extension study). Follow-up was completed 1422 women (N=49 were excluded either because they did not have a measurement of femoral neck BMD at baseline or no further data were available after the baseline visit). Participants were asked at each 6-month visit about fractures, and relevant radiographs or reports were reviewed. Nonpathological fractures and low-trauma fractures, defined as a fall from standing height or less or equivalent injury were included in analyses. Classification of osteoporotic fractures was specific to each calculator; FRAX-defined osteoporotic fractures were fractures of the shoulder, hip, or forearm and clinical vertebral fractures; Garvan-defined osteoporotic fractures were fractures of the hip, vertebrae (symptomatic), forearm, metacarpal, humerus, scapula, clavicle, distal femur, proximal tibia, patella, pelvis, or sternum. *Note: Data was only extracted for BMD of the femoral neck*
Predicted outcomes	Major osteoporotic fracture (Hip, shoulder, clinical vertebral [spine], forearm) Hip fracture
Discrimination outcomes	c-statistic/AUC

Study arms

- Major osteoporotic fracture (N = 229)
- No major osteoporotic fracture (N = 1193)
- Hip fracture (N = 57)
- No hip fracture (N = 1365)

Characteristics

Study-level characteristics

Characteristic	Study (N = 1422)
% Female No of events	n = 1422 ; % = 100
Mean age (SD) Mean (SD)	74.2 (4.2)
BMI (kg/m^2) Mean (SD)	26.5 (4.2)
Alcohol intake - Alcohol greater than or equal to 3 units per day No of events	n = 53 ; % = 3.7
Smoking status - Current smokers	n = 43 ; % = 3

Characteristic	Study (N = 1422)
No of events	
Previous fracture - Previous fracture during adult life	n = 476 ; % = 33.5
No of events	
Fall history - 0 falls in the past year	n = 1295 ; % = 91.1
No of events	
Fall history - 1 fall in the past year	n = 0 ; % = 0
No of events	
Fall history - 2 falls in the past year	n = 60 ; % = 4.2
No of events	
Fall history - 3 or more falls in the past year	n = 67 ; % = 4.7
No of events	
Rheumatoid arthritis or systemic lupus erythematosus - Rheumatoid arthritis	n = 2 ; % = 0.1
No of events	
Corticosteroid use - Glucocorticoid use	n = 0 ; % = 0
No of events	

Outcomes

1 **DXA BMD-FN t-score**

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 1193, N1 = 229	Hip fracture vs No hip fracture, N2 = 1365, N1 = 57
AUC Unadjusted Custom value	0.6 (95%CI 0.56-0.64)	0.64 (95%CI 0.57-0.72)

2 AUC - Polarity - Higher values are better

3 AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

4

5 **Critical appraisal - Critical Appraisal - QUADAS-2**

Overall risk of bias and directness	Risk of Bias	Outcome HF: Moderate *(For HF, study had an overall moderate risk of bias due to high risk in the domain flow and timing. This is because not all participants received the same reference standard (self-report during the extension study, however in the initial study self-reported fractures were confirmed by radiographs) and not all participants were included in the analyses (49 were excluded either because they did not have a measurement of femoral neck BMD at baseline or no further data were available after the baseline visit). All other domains were low risk.* Outcome MOF: High *(Study had overall high RoB due to above reason, and high risk of bias in the reference standard domain due to fracture ascertainment being limited to self-report.)*
Overall risk of bias and directness	Directness	Directly applicable *(Postmenopausal women were included and this group are likely to be suspected risk of fragility fracture)*

6

D.1.4 Boutroy, 2013

Bibliographic Reference	Boutroy, S; Hans, D; Sornay-Rendu, E; Vilayphiou, N; Winzenrieth, R; Chapurlat, R; 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; 2013; vol. 24 (no. 1); 77-85

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study included in review	OFELY cohort, see also Chapurlat 2020 which includes participants from this cohort.
Study type	Prospective cohort study Retrospective analysis of prospective cohort
Study location	France
Study setting	Health insurance company
Study dates	Participants were recruited in 1992–1993 and had scans between 2000 and 2001
Sources of funding	NR
Study sample	N=582 postmenopausal women (aged 31–89 years)

	On treatment at baseline: NR Received treatment during FU: 41%
Inclusion criteria	• postmenopausal Caucasian women • lumbar spine DXA scan between years 2000 and 2001
Exclusion criteria	NR
Bone assessment method	Dual X-ray absorptiometry BMD-LS T-score; BMD-total hip T-score
Bone assessment method - baseline	Areal bone mineral density (aBMD) was acquired at the lumbar spine (L1–L4) and total hip using a Hologic QDR 4500A DXA device (Hologic, Bedford, MA, USA) and measurements were performed by the same experienced operator. Spine scans were analysed with the APEX software version 12.7.3. T-scores were calculated based on the NHANES III reference curve for the total hip and Hologic reference curve for Caucasian women for the lumbar spine. Using the WHO classification, women were classified as normal (T-score>−1), osteopenic (−1≥T-score>−2.5), or osteoporotic (T-score≤−2.5) based on the lower values of their aBMD measurements at the lumbar spine or total hip. The lumbar spine trabecular bone score (LS_TBS) parameter was assessed on the same regions used for lumbar spine DXA scans, blinded from clinical outcome, using the TBS iNsight Software v1.7 (Med-Imaps, Bordeaux, France). LS_TBS was calculated as the mean value of individual measurements for vertebrae L1–L4.
Bone assessment method - length of follow up	Mean=7.8 years (SD 1.3) For 94 people with fracture, mean FU=8.0 years (SD 1.1)

	For 466 people with no fracture, mean FU=7.8 years (SD 1.4)
Bone assessment method - follow up	Reference standard: Clinical review confirmed by radiographs. Follow-up was completed in 560 women (N=22 were excluded because of severe scoliosis, poor-quality DXA scans, bone metastasis, Paget's disease, or BMI>35). Incident fractures were annually registered between the observation period and follow-up. Clinical vertebral fractures were collected every year. Vertebral fractures that did not reach clinical attention were assessed on the X-ray films performed every 4 years. Vertebral fractures were identified on lateral X-ray films of the thoracic and the lumbar spine by a trained rheumatologist blinded of lumbar spine aBMD and LS_TBS. Vertebral deformities due to other causes than osteoporosis such as osteoarthritis and Scheuermann's disease were excluded. Only low trauma fractures (such as those occurring after falls from standing height or less) were taken into account. All sites were included except the head, toes, and fingers. Prevalent fragility fractures were all those that occurred between the inclusion in the study (1992–1993) and the observation period, in addition to the fragility fractures of the wrist, humerus, vertebra, or hip that occurred before the inclusion in the study and after the age of 40 years. Fractured vertebrae were retrospectively excluded from the analysis. *Note: data not extracted for univariate analysis of TBS-lumbar spine*
Predicted outcomes	• Major osteoporotic fracture (Hip, shoulder, clinical vertebral [spine], forearm)
Discrimination outcomes	c-statistic/AUC

Study arms

- Major osteoporotic fracture (N = 81)
- No major osteoporotic fracture (N = 479)

Study-level characteristics

Characteristic	Study (N = 560)
% Female No of events	n = 560 ; % = 100
Mean age (SD) Mean (SD)	NA (NA)
Mean age (SD) - No fracture Mean (SD)	65.3 (7.6)
Mean age (SD) - Fracture Mean (SD)	70.4 (9.4)
BMI (kg/m^2) Mean (SD)	NA (NA)
BMI - No fracture Mean (SD)	24.6 (3.6)

Characteristic	Study (N = 560)
BMI - Fracture Mean (SD)	23.9 (3.7)

Outcomes

DXA BMD-LS t-score

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 479, N1 = 81
AUC - Unadjusted Custom value	0.62 (95%CI 0.55-0.68)
AUC - Adjusted for age, weight, and prevalent fracture Custom value	0.67 (95%CI 0.61-0.73)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

DXA BMD-TH t-score

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 479, N1 = 81
AUC - Unadjusted Custom value	0.68 (95%CI 0.62-0.74)
AUC - Adjusted for age, weight, and prevalent fracture Custom value	0.7 (95%CI 0.64-0.76)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	High *(High RoB due to concerns in the reference standard domain as only one person was involved in the interpretation of radiographs, and flow and timing since some exclusions.)*
Overall risk of bias and directness	Directness	Directly applicable *(Postmenopausal women were included and this group are likely to be suspected risk of fragility fracture)*

D.1.5 Briot, 2013

Bibliographic Reference	Briot, Karine; Paternotte, Simon; Kolta, Sami; Eastell, Richard; Felsenberg, Dieter; Reid, David M; Gluer, Claus-C; Roux, Christian; FRAX R: prediction of major osteoporotic fractures in women from the general population: the OPUS study.; PloS one; 2013; vol. 8 (no. 12); e83436

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	Prospective cohort study
Study location	Europe
Study setting	Community
Study dates	Recruitment between 1999 and 2001
Sources of funding	Sponsored by Eli Lilly, Sanofi-Aventis, Procter and Gamble Pharmaceuticals, Hoffman-La Roche, Pfizer, and Novartis
Study sample	N=1748 ambulatory European women aged >55 years (N=2409 women were included in the original OPUS Cohort study)

	On treatment at baseline or during FU: 60% Received treatment during FU: 8.9%
Inclusion criteria	• ambulatory European women aged > 55 years • with information on incident major osteoporotic fractures • enrolled in OPUS cohort
Exclusion criteria	• disorders precluding ultrasound and bone mineral density measurements • general and cognitive inability that precluded completing questionnaire
Population subgroups	Postmenopausal women who ever received anti-osteoporotic treatment at baseline (n=1050) Postmenopausal women who never received any anti-osteoporotic treatment at baseline (n=698)
Bone assessment method	Dual X-ray absorptiometry BMD-FN T-score, adjusted for age and fragility fracture history
Bone assessment method - baseline	Using standardised procedures and centralised quality control, participants had BMD of the lumbar spine and proximal femur measured using dual energy X-ray absorptiometry (DXA) in the postero-anterior projection (Hologic QDR-4500; Hologic, Bedford, MA, USA in Paris, Kiel and Sheffield centres) or in the antero-posterior (Lunar Expert devices; GE Lunar, Madison, USA in the Berlin and Aberdeen centres).
Bone assessment method - length of follow up	Mean=6.04 years (range 4.5-7.5)
Bone assessment method - follow up	Reference standard: Self-report with fracture confirmed by written radiographic or surgical reports. Follow-up was completed in 1748 participants. Incident fractures were self-reported and confirmed by written reports of radiographs or surgical reports. Radiographs at baseline and follow-up were preformed using the same procedures and

	evaluated centrally by two radiologists. A copy of the radiograph obtained by the participants physician was sent to the coordinating centre and compared with the baseline study radiograph. Major osteoporotic fractures were included, defined as a fracture of the hip, spine (clinical), wrist, or humerus. Pathologic fractures were excluded.
Predicted outcomes	• Major osteoporotic fracture (Hip, shoulder, clinical vertebral [spine], forearm)
Discrimination outcomes	c-statistic/AUC

Study arms

- Major osteoporotic fracture (N = 85)
- No major osteoporotic fracture (N = 1663)

Characteristics

Study-level characteristics

Characteristic	**Study (N = 1748)**
% Female No of events	n = 1748 ; % = 100
Mean age (SD) Mean (SD)	66.1 (6.8)

Characteristic	Study (N = 1748)
Mean age (SD) in subpopulation who never received any anti-osteoporotic treatment n=698 Mean (SD)	69.1 (6.4)
BMI (kg/m^2) Mean (SD)	26.7 (4.5)
BMI in subpopulation who never received any anti-osteoporotic treatment n=698 Mean (SD)	27.5 (4.6)
Alcohol intake No of events	n = NA ; % = NA
Alcohol intake - Alcohol intake greater than or equal to 3 units per day No of events	n = 11 ; % = 0.6
Alcohol intake - Alcohol intake greater than or equal to 3 units per day in subpopulation who never received any anti-osteoporotic treatment n=698 No of events	n = 3 ; % = 0.4
Smoking status - Current smoking No of events	n = 236 ; % = 13.8

Characteristic	Study (N = 1748)
Smoking status - Current smoking in subpopulation who never received any anti-osteoporotic treatment n=698 No of events	n = 80 ; % = 11.9
Previous fracture - History of low trauma fracture No of events	n = 742 ; % = 43.1
Previous fracture - History of low trauma fracture in subpopulation who never received any anti-osteoporotic treatment n=698 No of events	n = 286 ; % = 42.1
Rheumatoid arthritis or systemic lupus erythematosus - Rheumatoid arthritis No of events	n = 108 ; % = 6.4
Rheumatoid arthritis or systemic lupus erythematosus - Rheumatoid arthritis in subpopulation who never received any anti-osteoporotic treatment n=698 No of events	n = 48 ; % = 7.3
Type 1 diabetes No of events	n = 0 ; % = 0
Type 1 diabetes in subpopulation who never received any anti-osteoporotic treatment n=698	n = 0 ; % = 0

Characteristic	Study (N = 1748)
No of events	

Outcomes

DXA BMD-FN t-score

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 1663, N1 = 85
Unadjusted Custom value	0.65 (95%CI 0.58-0.71)
Adjusted for age Custom value	0.66 (95%CI 0.6-0.72)
Adjusted for age and fracture history Custom value	0.69 (95%CI 0.63-0.75)
No treatment Women who did not have anti-osteoporotic treatment before or during study (N=698; fractured=35, non-fractured=663) Custom value	0.58 (95%CI 0.47-0.69)

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 1663, N1 = 85
Treatment Women who had anti-osteoporotic treatment before or during study (N=1050; fractured=50, non-fractured=1000) Custom value	0.69 (95%CI 0.61-0.76)

AUC - Polarity - Higher values are better

Femoral neck

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	Moderate *(Study at moderate ROB due to concerns in the flow and timing domain. Radiographs of patients checked only if fracture was reported.)*
Overall risk of bias and directness	Directness	Directly applicable *(Population were postmenopausal women aged above 55 years and subsequently likely to have a suspected risk of fragility fracture)*

D.1.6 Center, 2004

Bibliographic Reference Center, Jacqueline R; Nguyen, Tuan V; Pocock, Nick A; Eisman, John A; 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; 2004; vol. 89 (no. 6); 2776-82

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study included in review	DOES cohort, see Chan 2013
Study type	Prospective cohort study
Study location	Dubbo, Australia
Study setting	Community
Study dates	1989 to February 2002
Sources of funding	Not reported
Study sample	N=1584 community-dwelling people (658 men/926 women). 19 participants (12 women, 7 men) who had hip fracture excluded because bone density assessment conducted more than 3 months after fracture. On treatment at baseline: NR Received treatment during FU: NR

Inclusion criteria	• Participating in Dubbo Osteoporosis Epidemiology Study • Age≥60 on 01/01/1989 obtained from electoral roll and other local approaches invited to participate in study • Informed consent
Exclusion criteria	• Traumatic hip fracture • Underlying conditions that could cause pathological fracture (for example, malignancy, metabolic bone disease)
Population subgroups	NA
Bone assessment method	Dual X-ray absorptiometry DXA: BMD-FN
Bone assessment method - baseline	Participants interviewed by nurse coordinator at initial and FU visits approximately every 2 years. Structured questionnaire used to collect data. BMD of femoral neck obtained by DXA (Lunar DPX-L densitometer, GE Lunar).
Bone assessment method - length of follow up	FU up to 13 years (1989 to 02/2002)
Bone assessment method - follow up	Reference standard: Low-trauma hip fractures identified by review of all radiology reports from the two radiology services serving Dubbo area, with circumstances obtained by personal interview after fracture. Most recent BMD measurement at femoral neck used if measure more than once before fracture. *Note: only data for hip fracture in women was extracted. Mean age of men in study was 69.2 years.*
Predicted outcomes	• Hip fracture

Discrimination outcomes	Sensitivity At T-score ≤-2.5 SD cut-off for areal BMD Specificity At T-score ≤-2.5 SD cut-off for areal BMD

Study arms

- Hip fracture in women (N = 73)
- No hip fracture in women (N = 853)
- Hip fracture in men (N = 23)
- No hip fracture in men (N = 635)

Characteristics

Study-level characteristics

Characteristic	Study (N = 1584)
% Female Sample size	n = 926; % = 58.5
Ethnicity White	n = 1584; % = 100

Characteristic	Study (N = 1584)
Sample size	

Arm-level characteristics

Characteristic	Hip fracture in women (N = 73)	No hip fracture in women (N = 853)	Hip fracture in men (N = 23)	No hip fracture in men (N = 635)
Mean age (SD) Mean age of women=70.6. Mean age of men=69.2 Mean (SD)	77 (7)	70 (7)	76 (8)	69 (6)

Outcomes

DXA BMD-FN t-score

Outcome	Hip fracture in women vs No hip fracture in women, N2 = 853, N1 = 73	Hip fracture in men vs No hip fracture in men, N2 = 635, N1 = 23
Sensitivity and specificity - Sensitivity At T-score <=-2.5SD areal BMD threshold Custom value	0.73 (95%CI 0.61-0.82)	0.43 (95%CI 0.23-0.66)

Outcome	Hip fracture in women vs No hip fracture in women, N2 = 853, N1 = 73	Hip fracture in men vs No hip fracture in men, N2 = 635, N1 = 23
Sensitivity and specificity - Specificity At T-score <=-2.5SD areal BMD threshold Custom value	0.81 (95%CI 0,78-0.84)	0.92 (95%CI 0.9-0.94)

Sensitivity and specificity - Polarity - Higher values are better

Hip fracture in men: TP=10, FP=51, FN=13, TN=584; Hip fracture in women: TP=53, FP=162, FN=20, TN=690. Raw data and 95%CIs estimated using reported sensitivity and specificity and RevMan 5.4.1 calculator.

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	Moderate *(Overall moderate risk of bias due to concerns about reference standard (unclear how many interpreters of X-rays there were) and flow and timing (not all participants had X-ray to confirm presence or absence of fracture))*
Overall risk of bias and directness	Directness	Directly applicable

D.1.7 Chan, 2013

Bibliographic Reference Chan, M Y; Nguyen, N D; Center, J R; Eisman, J A; Nguyen, T V; 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; 2013; vol. 24 (no. 3); 1015-22

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study included in review	DOES cohort, see Center 2004
Study type	Prospective cohort study
Study location	Dubbo, Australia
Study setting	Community
Study dates	1994-2011
Sources of funding	Partly supported by Australia National Health and Medical Research Council.
Study sample	N=702 adults (390 men, 312 women) with calcaneal QUS measurements On treatment at baseline: NR Received treatment during FU: NR
Inclusion criteria	• Participating in DOES population-based cohort study

	• Non-osteoporotic (BMD T-score >-2.5 SD at femoral neck (WHO definition))
Exclusion criteria	• Malignant or malignant bone disease • Traumatic fracture
Population subgroups	NA
Bone assessment method	Quantitative ultrasound QUS-heel: BUA, adjusted for age, falls and prior fracture (CUBA sonometer) Dual X-ray absorptiometry DXA: BMD-FN
Bone assessment method - baseline	Baseline characteristics obtained via use of structured questionnaire administered by trained nurse during initial and FU visits every 2 years. QUS measurements (BUA, mhz; Velocity of sound, m/sec) at calcaneus using CUBA sonometer (McCue Ultrasonics). Coefficients were 3.1 and 0.3%, respectively. BMD measured by DXA at femoral neck using GE Lunar DPX-L densitometer. Radiation dose <0.1 miuGy, coefficient of variation was 1.5%.
Bone assessment method - length of follow up	Median FU=12 years (range 0.1-17)
Bone assessment method - follow up	Reference standard: fragility fractures ascertained through X-ray reports from 2-3 radiology centres in Dubbo. Vertebral fractures clinically diagnosed with no systemic X-ray screening for asymptomatic vertebral fractures. *Note: only data for hip fracture in women was extracted. Mean age of men in study was 72.4 years.*
Predicted outcomes	• Hip fracture

Discrimination outcomes	c-statistic/AUC

Study arms

- Major osteoporotic fracture in women (N = 80)
- No major osteoporotic fracture in women (N = 232)
- Hip fracture in women (N = 12)
- No hip fracture in women (N = 300)

Characteristics

Study-level characteristics

Characteristic	Study (N = 702)
% Female	n = 312 ; % = 44.4
Sample size	

Arm-level characteristics

Characteristic	Major osteoporotic fracture in women (N = 80)	No major osteoporotic fracture in women (N = 232)	Hip fracture in women (N = 12)	No hip fracture in women (N = 300)
Mean age (SD) Mean (SD)	72 (5)	70 (4.9)	75 (5.5)	NR (NR)
BMI (kg/m2) Mean (SD)	26.6 (4.1)	27 (4.5)	26 (3.1)	NR (NR)
Smoking status History of smoking **Sample size**	n = 25 ; % = 31	n = 79 ; % = 34	n = 5 ; % = 42	n = NR ; % = NR
Previous fracture Previous fracture after age 50 **Sample size**	n = 49 ; % = 21	n = 23 ; % = 10	n = 2 ; % = 17	n = NR ; % = NR
Fall history Falls in last 12 months Sample size	n = 33; % = 41	n = 56; % = 24	n = 5 ; % = 42	n = NR ; % = NR

Outcomes

DXA BMD-FN t-score

Outcome	Major osteoporotic fracture in women vs No major osteoporotic fracture in women, N2 = 232, N1 = 80	Hip fracture in women vs No hip fracture in women, N2 = 300, N1 = 12
AUC Adjusted for age, falls, and prior fracture Custom value	0.68 (95%CI 0.61-0.75)	0.77 (95%CI 0.61-0.92)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

QUS-heel BUA

Outcome	Major osteoporotic fracture in women vs No major osteoporotic fracture in women, N2 = 232, N1 = 80	Hip fracture in women vs No hip fracture in women, N2 = 300, N1 = 12
AUC Adjusted for age, falls, and prior fracture Custom value	0.71 (95%CI 0.64-0.78)	0.85 (95%CI 0.75-0.95)

AUC - Polarity - Higher values are better

Cuba sonometer

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	High *(Overall high risk of bias due to concerns about patient selection (excludes osteoporotic patients), reference standard (unclear how many interpreters of X-rays there were), and flow and timing (not all participants had X-ray to confirm presence or absence of fracture))*
Overall risk of bias and directness	Directness	Directly applicable for HF Partially applicable for MOF *(Reference standard includes all fragility fractures so may include more fractures than those defined by MOF.)*

D.1.8 Chapurlat, 2020

Bibliographic Reference Chapurlat, Roland; Bui, Minh; Sornay-Rendu, Elisabeth; Zebaze, Roger; Delmas, Pierre D; Liew, Danny; Lespessailles, Eric; Seeman, Ego; 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; 2020; vol. 35 (no. 5); 833-844

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study included in review	Includes participants from OFELY cohort, see Boutroy 2013.

Study type	Prospective cohort study
Study location	France
Study setting	Not reported
Study dates	The OFELY study commenced in 1992, and this analyses focussed on years 2006–2008. QUALYOR cohort recruited from 2010-2013.
Sources of funding	Not reported
Study sample	N=2128 women (589 postmenopausal women from the OFELY study and 1539 women from the QUALYOR study) On treatment at baseline: NR Received treatment during FU: NR
Inclusion criteria	• Postmenopausal women from the OFELY cohort study • Women with T-scores at the hip or spine between −1.0 and −2.5 SD with clinical risk factors for fracture or −3.0 SD without clinical risk factors from the QUALYOR cohort study
Exclusion criteria	Not reported
Bone assessment method	Dual X-ray absorptiometry BMD-FN T-score
Bone assessment method - baseline	Femoral neck BMD was measured using Hologic Discovery A in QUALYOR and QDR 4500 in OFELY. Distal radial images obtained using HR-pQCT (Xtreme CT). *Note: Study examines discriminatory power of HR-pQCT using structural fragility score (SFS). Data for this was not extracted.*

Bone assessment method - length of follow up	OFELY cohort, median FU=9.4 years QUALYOR cohort: 5 years
Bone assessment method - follow up	Reference standard: Radiographs, DXA VFA, or clinical reports. Follow-up was completed in 2100 women (N=28 were excluded due to missing values for BMD, FRAX, or SFS). Fractures of the head, toes, and fingers were excluded. *Note: Data extracted for 8-year follow up for MOF only for whole sample. Data for subgroups (osteopenia/normal BMD; osteoporosis; >70 years with osteopenia/normal BMD), all fractures and 2 and 4 year follow up not extracted. Data for sensitivity/specificity for vBMD measured by HR-QCT distal radius not extracted as raw data was not extractable.*
Predicted outcomes	• Major fragility fracture (Hip, shoulder, clinical vertebral [spine], forearm)
Discrimination outcomes	c-statistic/AUC

Study arms

- Major osteoporotic fracture (N = 126)
- No major osteoporotic fracture (N = 1413)

Characteristics

Study-level characteristics

Characteristic	Study (N = 2100)
% Female No of events	n = 2100 ; % = 100
Mean age (SD) - OFELY cohort Mean (SD)	68 (NA)
Mean age (SD) - QUALYOR cohort Mean (SD)	65.9 (NA)

Outcomes

DXA BMD-FN t-score

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 1413, N1 = 126
AUC Unadjusted Custom value	0.628 (95%CI (0.56-0.7)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	High *(Study had an overall high risk of bias due to high risk in the three domains patient selection, reference standard and flow and timing. This is because it is unclear whether a random or consecutive sample of participants were enrolled and the exclusion criteria is not reported. Multiple reference standards were used (radiographs, DXA-VFA, or clinical reports) and no details about interpretation of radiographs provided. However, knowledge of the index test would be unlikely to introduce bias for the outcome (fracture). Multiple reference standards were used (radiographs, DXA-VFA, or clinical reports) so not all patients received same test. Not all participants were included in the analysis (28 were excluded due to missing values for BMD, FRAX, or SFS).)*
Overall risk of bias and directness	Directness	Directly applicable *(Postmenopausal women were included and this group are likely to be with suspected risk of fragility fracture)*

D.1.9 Crandall, 2020

Bibliographic Reference Crandall, C.J.; Larson, J.; Wright, N.C.; Laddu, D.; Stefanick, M.L.; Kaunitz, A.M.; Watts, N.B.; Wactawski-Wende, J.; Womack, C.R.; Johnson, K.C.; Carbone, L.D.; Jackson, R.D.; Ensrud, K.E.; Serial Bone Density Measurement and Incident Fracture Risk Discrimination in Postmenopausal Women; JAMA Internal Medicine; 2020; vol. 180 (no. 9); 1232-1240

Study details

Secondary publication of another included study- see primary study for details	NA
Study type	Prospective cohort study

Study location	USA
Study setting	Secondary (Clinical centres)
Study dates	Follow-up between 1993 and 2010 and data analysis between May 2019 and December 2019
Sources of funding	National Heart, Lung, and Blood, Institute, National Institutes of Health, US Department of Health, and Human Services
Study sample	N=161909 postmenopausal women aged 50 to 79 years were recruited into the original WHI study population This analysis included N=7419 participants On treatment at baseline: 47% on HRT Received treatment during FU: NR
Inclusion criteria	• postmenopausal women • aged 50 to 79 years • free from serious cardiac, pulmonary, renal, and hepatic conditions
Exclusion criteria	Not reported
Bone assessment method	Dual X-ray absorptiometry BMD-FN; BMD-LS; BMD-total hip
Bone assessment method - baseline	Participants had BMD and appendicular lean mass (sum of lean mass of the arms and legs) measured using dual-energy x-ray absorptiometry (DXA) with Hologic QDR2000 or 4500W machines (Hologic, Inc) at WHI study baseline and follow-up year 3. The DXA quality assurance procedures included cross clinic use of hip and spine phantom scans, further evaluation of scans with specific problems, and review of a random sample of all scans.

Bone assessment method - length of follow up	Mean=12.1 years (SD 3.4) Mean=9.0 years (SD 3.5) years after second BMD measurement
Bone assessment method - follow up	Reference standard: Self-report with fracture adjudicated using medical records. Follow-up was completed in 7419 participants (Of the 9304 participants who underwent both baseline and year 3 BMD measurements, data was excluded from 1885 participants due to the following reasons; use of bisphosphonates, calcitonin, selective oestrogen receptor modulators, or a combination of those medications prior to their year 3 BMD measurement; non-attendance of follow-up visits following their year 3 BMD measurement; a history of MOF at study baseline or between BMD measurements; missing covariate data (regarding hormone use, history of fracture, or BMI). Fractures were self-reported on annual questionnaires, and self-reported hip fractures were adjudicated using medical records. The validity of self-reporting of fractures obtained in WHI was reported as good; hip (78%), forearm/wrist (81%) fractures, and clinical vertebral [spine] fractures (51%). *Note: Data only extracted for baseline DXA BMD measurements.*
Predicted outcomes	• Major osteoporotic fracture (Hip, spine, radius, ulna, wrist, upper arm, or shoulder) • Hip fracture
Discrimination outcomes	c-statistic/AUC
Additional comments	The study aims to assess whether a second BMD measurement approximately 3 years after the initial assessment is associated with improved ability to estimate fracture risk beyond the baseline BMD measurement alone, therefore Baseline BMD, BMD change, and Baseline BMD + BMD change are all reported.

Study arms

- Major osteoporotic fracture (N = 732)

- No major osteoporotic fracture (N = 6687)
- Hip fracture (N = 127)
- No hip fracture (N = 7292)

Characteristics

Study-level characteristics

Characteristic	Study (N = 7419)
% Female No of events	n = 7419 ; % = 100
Mean age (SD) Mean (SD)	66.1 (7.2)
Ethnicity - white No of events	n = 5699 ; % = 77
Ethnicity - African American No of events	n = 1116 ; % = 15
Ethnicity - Hispanic No of events	n = 434 ; % = 6
Ethnicity - Other/unknown	n = 170 ; % = 2

Characteristic	Study (N = 7419)
No of events	
BMI (kg/m^2) Mean (SD)	28.7 (6)
Smoking status - Current smoker No of events	n = 501 ; % = 7
Fall history - 0 falls in the last year No of events	n = 5176 ; % = 70
Fall history - >0 to <2 falls in the last year No of events	n = 1729 ; % = 23
Fall history - Greater than or equal to 2 falls in the last year No of events	n = 347 ; % = 5
Antidepressant use - Selective serotonin reuptake inhibitor No of events	n = 527 ; % = 7
Corticosteroid use - Systemic corticosteroid No of events	n = 17 ; % = 0.22
Hormone replacement therapy - Current oestrogen therapy use (oral or transdermal)	n = 3454 ; % = 47

Characteristic	Study (N = 7419)
No of events	

Outcomes

DXA BMD-FN t-score

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 6687, N1 = 732	Hip fracture vs No hip fracture, N2 = 7292, N1 = 127
AUC Adjusted for hormone use and clinic, age, race/ethnicity, fracture history, physical activity, physical function, and fall history. Custom value	0.61 (95%CI 0.59-0.63)	0.71 (95%CI 0.67-0.75)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

DXA BMD-LS t-score

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 6687, N1 = 732	Hip fracture vs No hip fracture, N2 = 7292, N1 = 127
AUC Adjusted for hormone use and clinic, age, race/ethnicity,	0.59 (95%CI 0.57-0.61)	0.6 (95%CI 0.56-0.65)

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 6687, N1 = 732	Hip fracture vs No hip fracture, N2 = 7292, N1 = 127
fracture history, physical activity, physical function, and fall history. Custom value		

AUC - Polarity - Higher values are better

Lumbar spine vertebrae scanned not reported

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

DXA BMD-TH t-score

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 6687, N1 = 732	Hip fracture vs No hip fracture, N2 = 7292, N1 = 127
AUC Adjusted for hormone use and clinic Custom value	0.61 (95%CI 0.59-0.63)	0.71 (95%CI 0.67-0.75)

AUC - Polarity - Higher values are better

Total hip

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	High *(Study had an overall high risk of bias due to high risk in the two domains patient selection and flow and timing. This is because It is unclear whether a random or consecutive sample of participants were enrolled, and the exclusion criteria is not reported. Not all participants received the same reference standard as only self-reported fractures were verified by medical records. Not all participants were included in the analysis (1885 were excluded due to the use of bisphosphonates, calcitonin, selective oestrogen receptor modulators, or a combination of those medications prior to their year 3 BMD measurement, nonattendance of follow-up visits following their year 3 BMD measurement, a history of MOF at study baseline or between BMD measurements or missing covariate data). For MOF, there is also high risk of bias for the reference standard domain as no information about whether radiographs used was provided.)*
Overall risk of bias and directness	Directness	Directly applicable

D.1.10 Cummings, 1994

Bibliographic Reference Cummings, S R; Marcus, R; Palermo, L; Ensrud, K E; Genant, H K; 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; 1994; vol. 9 (no. 9); 1429-32

Study details

Secondary publication of another included study- see primary study for details	NA

Other publications associated with this study included in review	Participants from SOF cohort included in Ensrud 2024 and Hillier 2007.
Study type	Prospective cohort study
Study location	USA
Study setting	Community
Study dates	Participants were enrolled during September 1986 to October 1988, and a second examination of the cohort between 1988 and 1990
Sources of funding	Supported by grants from the US Public Health Service, National Institutes of Health, and the Research Service of the US Department of Veterans Affairs
Study sample	N=9704 white women aged ≥ 65 years were enrolled in the Study of Osteoporotic Fractures N=7963 of these were included in the analysis On treatment at baseline: NR Received treatment during FU: NR
Inclusion criteria	• white women • aged ≥65 years • no previous hip fractures
Exclusion criteria	Not reported

Bone assessment method	Dual X-ray absorptiometry BMD-FN, bone mineral apparent density (BMAD) of the femoral neck
Bone assessment method - baseline	Bone mass and dimensions of the femoral neck were measured using QDR 1000 densitometers (Hologic, Inc., Waltham MA) The mean coefficient of variation between centres was 1.2% for the femoral neck for two research staff who visited all centres and 0.8% for the DXA scans with a 1.0g/cm2. Bone mineral density was calculated by the Hologic QDR 1000 system as bone mineral content/projected area Ap of the femoral neck, expressed as mg/cm2. Bone mineral apparent density was estimated as BMC/(Ap x mean width) expressed in g/cm3. The default length of bone scanned (15cm) was used and width calculated as Ap/length, thus BMAS=BMC/length x width 2 which is equivalent to BMD/width.
Bone assessment method - length of follow up	Mean=2.1 years after second examination
Bone assessment method - follow up	Reference standard: Self-report with fracture confirmed by radiologist using radiographic reports or preoperative radiographs. Follow-up was completed in 7963 participants. Participants were contacted about fractures every 4 months by letter or phone. All confirmed hip fractures classified into either cervical or intertrochanteric. *Note: Data not extracted for BMAD and BMC measures*
Predicted outcomes	• Hip fractures
Discrimination outcomes	c-statistic/AUC

Study arms

- Hip fracture (N = 83)
- No hip fracture (N = 7880)

Characteristics

Study-level characteristics

Characteristic	Study (N = 7963)
% Female No of events	n = 7963; % = 100
Mean age (SD) Mean (SD)	73.2 (NA)
Ethnicity No of events	n = NA; % = NA
Ethnicity - Race, white No of events	n = 7941; % = 99.7

Outcomes

DXA BMD-FN raw scores

Outcome	Hip fracture vs No hip fracture, N2 = 7880, N1 = 83
AUC Raw scores Custom value	0.73 (95%CI 0.673-0.781)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group. 95%CIs calculated using equations in Section 3.1, Debray 2018.

DXA BMD-TH raw scores

Outcome	Hip fracture vs No hip fracture, N2 = 7880, N1 = 83
AUC Custom value	0.76 (95%CI 0.705-0.808)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group. 95%CIs calculated using equations in Section 3.1, Debray 2018.

DXA BMD-trochanter raw scores

Outcome	Hip fracture vs No hip fracture, N2 = 7880, N1 = 73
AUC Custom value	0.78 (95%CI 0.726-0.826)

AUC - Polarity - Higher values are better

1 AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group. 95%CIs calculated using equations in Section 3.1, Debray 2018.

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	High *(Study had an overall high risk of bias due to concerns in the domains patient selection and flow and timing. This is because it is unclear whether a random or consecutive sample of participants were enrolled, and the exclusion criteria is not reported. Furthermore, not all participants received the same reference standard as only self-reported fractures were confirmed by radiographic evidence.)*
Overall risk of bias and directness	Directness	Directly applicable

D.1.11 Dargent-Molina, 2003

Bibliographic Reference	Dargent-Molina, P; Piault, S; Breart, G; 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; 2003; vol. 14 (no. 12); 969-77

Study details

Secondary publication of another included study- see primary study for details	NA

Other publications associated with this study included in review	EPIDOS cohort, see Dargent-Molina 1999 and Hans 2004.
Study type	Prospective cohort study
Study location	France
Study setting	Community (participants were recruited through mailings using large population-based listings such as voter-registration rolls)
Study dates	Not reported
Sources of funding	Supported by a contract INSERM-MSD-Chibret
Study sample	N=7575 women aged ≥75 years were recruited into the EPIDOS study N=5910 women were included in this analysis On treatment at baseline: NR Received treatment during FU: NR
Inclusion criteria	• French women • aged ≥75 years
Exclusion criteria	• participants who did not have both BUA and BMD measurements
Bone assessment method	Quantitative ultrasound QUS-heel-BUA using Achilles sonometer

	Dual X-ray absorptiometry BMD-FN, adjusted for age
Bone assessment method - baseline	Participants completed a structured questionnaire, a clinical and functional examination, and bone investigations at baseline. Femoral neck BMD was measured by dual-energy X-ray absorptiometry (DXA) with a Lunar DPX-Plus. Ultrasound measurements of bone fragility were performed at the calcaneus with the Lunar Achilles ultrasound system.
Bone assessment method - length of follow up	Mean=3.7 years (SD 0.8), equal to 21,732 woman-years
Bone assessment method - follow up	Reference standard: Self-report. Follow-up for incident fracture was every 4 months and was completed in 5910 women who had both BUA and BMD measurements (The Achilles device was not available to some centres at the beginning of the EPIDOS recruitment meaning BUS data was not available for all participants).
Predicted outcomes	• Hip fracture
Discrimination outcomes	Sensitivity At fracture risk greater than 20 per 1000 woman-years Specificity At fracture risk greater than 20 per 1000 woman-years
Duration of follow-up	At least 4 years

Study arms

- Hip fracture (N = 231)
- No hip fracture (N = 5679)

Characteristics

Study-level characteristics

Characteristic	Study (N = 5910)
% Female No of events	n = 5910; % = 100
Mean age (SD) Mean (SD)	80.5 (3.8)

Outcomes

DXA BMD-FN

Outcome	Hip fracture vs No hip fracture, N2 = 5679, N1 = 231
Sensitivity at BMD threshold below which each ventile has fracture risk>20 per 1000 woman-years Custom value	0.35 (95%CI 0.29-0.41)
Specificity at given sensitivity	0.86 (95%CI 0.85-0.87)

Outcome	Hip fracture vs No hip fracture, N2 = 5679, N1 = 231
Custom value	

Sensitivity at threshold below which each ventile has risk>20 per 1000 woman-years - Polarity - Higher values are better

Specificity at given sensitivity - Polarity - Higher values are better

TP=81, FP=795, FN=150, TN=4884. Raw data estimated using reported sensitivity and specificity and RevMan 5.4.1 calculator.

QUS-heel BUA Achilles

Outcome	Hip fracture vs No hip fracture, N2 = 5679, N1 = 231
Sensitivity at BUA threshold below which each ventile has fracture risk>20 per 1000 woman-years Custom value	0.15 (95% CI 0.11-0.2)
Specificity at given sensitivity Custom value	0.95 (95%CI 0.95-0.96)

Sensitivity at threshold below which each ventile has risk>20 per 1000 woman-years - Polarity - Higher values are better

Specificity at given sensitivity - Polarity - Higher values are better

Achilles sonometer

TP=35, FP=284, FN=196, TN=5395. Raw data estimated using reported sensitivity and specificity and RevMan 5.4.1 calculator.

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	Moderate *(Study had an overall moderate risk of bias due to high risk in the domain reference standard. This is because there was a reliance on the self-reporting of fractures as this was the only reference standard used. It is unclear whether the reference standard was interpreted without knowledge of the results of the index test, however knowledge of the index test would be unlikely to introduce bias for the outcome (fracture). All other domains were low risk.)*
Overall risk of bias and directness	Directness	Directly applicable

D.1.12 Dargent-Molina, 1999

Bibliographic Reference	Dargent-Molina, P; Schott, A M; Hans, D; Favier, F; Grandjean, H; Baudoin, C; Meunier, P J; Breart, G; 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; 1999; vol. 9 (no. 2); 188-92

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with	EPIDOS cohort, see Dargent-Molina 2003 and Hans 2004

this study included in review	
Study type	Prospective cohort study
Study location	France
Study setting	Community
Study dates	January 1992 and January 1994
Sources of funding	Supported by INSERM-MSD-Chibre
Study sample	N=7575 women aged ≥75 years were recruited into the EPIDOS study and of these, 5978 had BUA measured (ultrasound machines were not available at the beginning of the recruitment process) This analysis included 5895 women who had no missing values for BMD, BUA, or gait speed. On treatment at baseline: NR Received treatment during FU: NR
Inclusion criteria	• women aged ≥75 years • with no missing values for BMD, BUA, or gait speed
Exclusion criteria	Not reported
Bone assessment method	Quantitative ultrasound QUS-heel-BUA using Achilles sonometer Dual X-ray absorptiometry

	BMD-FN
Bone assessment method - baseline	BMD was measured at the femoral neck by dual-energy X-ray absorptiometry (DXA) with a Lunar DPX-Plus, and BUA was measured at the calcaneus with a Lunar Achilles device (Lunar Corporation, Madison, WI).
Bone assessment method - length of follow up	Mean (SD)=2.75 years (SD 0.7)
Bone assessment method - follow up	Reference standard: Self-report. Participants received a mail questionnaire every 4 months, for reporting incident fractures and follow-up was completed in 5895 women (206 refused to continue participating,15 were lost to follow-up and 340 women died). The study reported data under three different definitions for the high-risk group: the top 50% (50% fracture risk), 25% (75% fracture risk) and 10% (90% fracture risk). The 50% cut-off was chosen on the basis of the WHO definition of osteoporosis (BMD more than 2.5 SD below the young adult mean) and approximately 50% of the EPIDOS population was osteoporotic at baseline.
Predicted outcomes	• Hip fracture
Discrimination outcomes	Sensitivity Sensitivity at 10%, 25%, and 50% fracture risk threshold
Duration of follow-up	

Additional comments	The study reported data under three different definitions for the high-risk group: the 50%, 25% and10% of women who are at highest risk. The 50% cut-off was chosen on the basis of the WHO definition of osteoporosis (BMD more than 2.5 SD below the young adult mean) and approximately 50% of the EPIDOS population was osteoporotic at baseline. The 25% and 10% cut-offs were also reported in the context of mass screening. In this study, sensitivity corresponds to the proportion of women with hip fracture who had been classified as being at high risk at baseline, and specificity corresponds to the proportion of women without hip fracture who had not been classified as being at high risk at baseline

Study arms

- Hip fracture (N = 170)
- No hip fracture (N = 5725)

Characteristics

Study-level characteristics

Characteristic	Study (N = 5895)
% Female No of events	n = 5895; % = 100
Mean age (SD) - Hip fracture cases Mean (SD)	82.6 (4.5)

Characteristic	Study (N = 5895)
Mean age (SD) - Non hip fracture cases Mean (SD)	80.4 (3.7)

Outcomes

DXA BMD-FN

Outcome	Hip fracture vs No hip fracture, N2 = 5725, N1 = 170
Sensitivity Custom value	NA
Top 50% highest risk group (50% fracture risk) - unadjusted Custom value	0.78
Top 50% highest risk group (50% fracture risk) - BMD adjusted for age TP=145, FN=25. Custom value	0.85 (95%CI 0.79-0.90)
Top 25% highest risk group (75% fracture risk) - unadjusted Custom value	0.49
Top 25% highest risk group (75% fracture risk) - BMD adjusted for age TP=88, FN=82. Custom value	0.52 (95%CI 0.44-0.59)

Outcome	Hip fracture vs No hip fracture, N2 = 5725, N1 = 170
Top 10% highest risk group (90% fracture risk) - unadjusted Custom value	0.25
Top 10% highest risk group (90% fracture risk) - BMD adjusted for age TP=49, FN=121. Custom value	0.29 (95%CI 0.22-0.36)

Sensitivity - Polarity - Higher values are better

Raw data and 95%CIs estimated using reported sensitivity and specificity and RevMan 5.4.1 calculator.

QUS-heel BUA Achilles

Outcome	Hip fracture vs No hip fracture, N2 = 5725, N1 = 170
Sensitivity Custom value	NA
Top 50% highest risk group (50% fracture risk) - unadjusted Custom value	0.69
Top 50% highest risk group (50% fracture risk) - adjusted for age TP=126, FN=44 Custom value	0.74 (95%CI 0.67-0.81)
Top 25% highest risk group (75% fracture risk)- unadjusted	0.43

Outcome	Hip fracture vs No hip fracture, N2 = 5725, N1 = 170
Custom value	
Top 25% highest risk group (75% fracture risk) - adjusted for age TP=94, FN=76 Custom value	0.55 (95%CI 0.47-0.63)
Top 10% highest risk group (90% fracture risk)- unadjusted Custom value	0.25
Top 10% highest risk group (90% fracture risk) - adjusted for age TP=49, FN=121 Custom value	0.29 (95%CI 0.22-0.36)

Sensitivity - Polarity - Higher values are better

Achilles sonometer

Raw data and 95%CIs estimated using reported sensitivity and specificity and RevMan 5.4.1 calculator.

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	High *(Study had an overall high risk of bias due to high risk in the three domains patient selection, reference standard and flow and timing. This is because the exclusion criteria was not reported. There was a reliance on the self-reporting of fractures as this was the only reference standard used. It is unclear whether the*

		reference standard was interpreted without knowledge of the results of the index test, however knowledge of the index test would be unlikely to introduce bias for the outcome (fracture). Not all participants completed follow-up and subsequently were not included in the analysis (3206 refused to continue participating,15 were lost to follow-up and 340 died).)
Overall risk of bias and directness	Directness	Directly applicable

D.1.13 Ensrud, 2024

Bibliographic Reference Ensrud, Kristine E; Schousboe, John T; Crandall, Carolyn J; Leslie, William D; Fink, Howard A; Cawthon, Peggy M; Kado, Deborah M; Lane, Nancy E; Cauley, Jane A; Langsetmo, Lisa; Hip Fracture Risk Assessment Tools for Adults Aged 80 Years and Older.; JAMA network open; 2024; vol. 7 (no. 6); e2418612

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study included in review	Includes data from 3 cohorts (SOF, Health ABC, and MrOs). For SOF cohort, see Cummings 1994 and Hillier 2007. For MrOS cohort, see Bauer 2007, Black 2008, Gourlay 2017, Kwok 2012, and Sheu 2011.
Study type	Prospective cohort study

Study location	USA
Study setting	Community
Study dates	Index examination between 1997 to 2016, and data analysis between March 2023 to April 2024
Sources of funding	The study was supported by the National Institute on Aging (NIA). The SOF and the MrOS studies were supported by NIH funding, and the HealthABC study was supported by the NIA. The following institutes also provided support: the National Institute of Arthritis and Musculoskeletal and Skin Diseases, the National Center for Advancing Translational Sciences, and NIH Roadmap for Medical Research
Study sample	N=19344 women from the three cohorts (SOF, MrOs and HealthABC) were eligible for inclusion (N=10366 from SOF, N=5994 from MrOs and N=3075 from HealthABC) N=8890 adults aged ≥80 years who had hip BMD were included in this analysis (N=4101 from SOF, N=3205 from MrOs and N=1584 from HealthABC) On treatment at baseline: 10.2% women, 4.4% men Received treatment during FU: NR
Inclusion criteria	• Adults aged ≥80 years • with hip BMD • from the SOF, MrOs or HealthABC cohorts *Inclusion criteria for SOF cohort* • Women aged≥65 years • Able to walk without assistance from another person • Absence of bilateral hip replacement *Inclusion criteria for MrOS cohort* • Men aged≥65 years

	• Able to walk without assistance from another person • Absence of bilateral hip replacement *Inclusion criteria for Health ABC cohort* • Adults aged 70-79 years • No self-reported mobility difficulty or disability
Exclusion criteria	Not reported
Bone assessment method	Dual X-ray absorptiometry BMD-FN T-score
Bone assessment method - baseline	Bone assessment methods are not described. In-person index examination including FNBMD measurement and clinical risk factor assessment were at years 10 or 16 of SOF; years 3, 5 or 6, 8, or 10 for Health ABC; and years 5, 7, or 14 for MrOS
Bone assessment method - length of follow up	Mean=4.4 years (SD 1.2)
Bone assessment method - follow up	Reference standard: Self-report with fracture confirmed by radiology reports. Participants were contacted every 4 months (SOF and MrOS) or 6 months (HealthABC) and followed to a maximum of 5 years until an event (hip fracture or death) or censoring. Over 95% of contacts in each cohort were completed in active participants. Follow up was completed in 8890 participants
Predicted outcomes	• Hip fracture for men • Hip fracture for women

Discrimination outcomes	c-statistic/AUC

Study arms

- Hip fracture in men (N = 123)
- No hip fracture in men (N = 3861)
- Hip fracture in women (N = 321)
- No hip fracture in women (N = 4585)

Characteristics

Study-level characteristics

Characteristic	Study (N = 8890)
% Female No of events	n = 4906 ; % = 55.2
Mean age (SD) - Women Mean (SD)	82.6 (2.7)
Mean age (SD) - Men	82.7 (2.7)

Characteristic	Study (N = 8890)
Mean (SD)	
Race and ethnicity - Black (women) No of events	n = 545 ; % = 11.1
Race and ethnicity - White (women) No of events	n = 4361 ; % = 88.9
Race and ethnicity - Other (women) No of events	n = 0 ; % = 0
Ethnicity - Race and ethnicity - Black (men) No of events	n = 321 ; % = 8.1
Ethnicity - Race and ethnicity - White (men) No of events	n = 3475 ; % = 87.2
Ethnicity - Race and ethnicity - Other (men) No of events	n = 188 ; % = 4.7
BMI (kg/m^2) - women Mean (SD)	26.6 (4.8)
BMI (kg/m^2) - Men	26.8 (3.8)

Characteristic	Study (N = 8890)
Mean (SD)	
Alcohol intake - greater than or equal to 3 drinks per day (women) No of events	n = 38 ; % = 0.8
Alcohol intake - greater than or equal to 3 drinks per day (men) No of events	n = 145 ; % = 3.6
Smoking status - current smoker (women) No of events	n = 152 ; % = 3.1
Smoking status - current smoker (men) No of events	n = 74 ; % = 1.9
Previous fracture - 0 fractures since age 50 years (women) No of events	n = 2413 ; % = 49.2
Previous fracture - 1 fracture since age 50 years (women) No of events	n = 974 ; % = 19.9
Previous fracture - 2 fractures since age 50 years (women) No of events	n = 950 ; % = 19.4
Previous fracture - 3 or more fractures since age 50 years (women)	n = 569 ; % = 11.6

Characteristic	Study (N = 8890)
No of events	
Previous fracture - 0 fractures since age 50 years (men)	n = 2849 ; % = 71.5
No of events	
Previous fracture - 1 fracture since age 50 years (men)	n = 791 ; % = 19.9
No of events	
Previous fracture - 2 fractures since age 50 years (men)	n = 214 ; % = 5.4
No of events	
Previous fracture - 3 or more fractures since age 50 years (men)	n = 130 ; % = 3.3
No of events	
Fall history - 0 falls in past year (women)	n = 3257 ; % = 66.4
No of events	
Fall history - 1 fall in past year (women)	n = 956 ; % = 19.5
No of events	
Fall history - 2 falls in past year (women)	n = 391 ; % = 8
No of events	
Fall history - 3 or more falls in past year (women)	n = 302; % = 6.2

Characteristic	Study (N = 8890)
No of events	
Fall history - 0 falls in past year (men) No of events	n = 2627; % = 65.9
Fall history - 1 fall in past year (men) No of events	n = 703; % = 17.6
Fall history - 2 falls in past year (men) No of events	n = 499; % = 12.5
Fall history - 3 or more falls in past year (men) No of events	n = 155; % = 3.9
Rheumatoid arthritis or systemic lupus erythematosus - Rheumatoid arthritis (women) No of events	n = 597; % = 12.2
Rheumatoid arthritis or systemic lupus erythematosus - Rheumatoid arthritis (men) No of events	n = 213; % = 5.3
Corticosteroid use No of events	n = NA ; % = NA
Corticosteroid use - Oral glucocorticoid use (women)	n = 165 ; % = 3.4

Characteristic	Study (N = 8890)
No of events	
Corticosteroid use - Oral glucocorticoid use (men) No of events	n = 108 ; % = 2.7

Outcomes

DXA BMD-FN t-score

Outcome	Hip fracture in men vs No hip fracture in men, N2 = 3861, N1 = 123	Hip fracture in women vs No hip fracture in women, N2 = 4585, N1 = 321
AUC Unadjusted Custom value	0.77 (95%CI 0.73-0.81)	0.72 (95%CI 0.69-0.75)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	High *(Study had an overall high risk of bias due to high risk in the two domains patient selection and flow and timing. This is because it is unclear whether a random or consecutive sample of participants were enrolled, and the exclusion criteria is not reported. Not all participants received the same reference standard as only*

		self reported fractures were verified by radiographic reports. Not all participants were included in the analysis.)
Overall risk of bias and directness	Directness	Directly applicable

D.1.14 Hans, 2004

Bibliographic Reference	Hans, D; Schott, A M; Duboeuf, F; Durosier, C; Meunier, P J; Does follow-up duration influence the ultrasound and DXA prediction of hip fracture? The EPIDOS prospective study.; Bone; 2004; vol. 35 (no. 2); 357-63

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study included in review	EPIDOS cohort, see Dargent-Molina 1999/2003
Study type	Prospective cohort study
Study location	France

Study setting	Community
Study dates	January 1992 to January 1994
Sources of funding	The EPIDOS study was supported by an INSERM/MSD-Chibret contract
Study sample	N=7598 Caucasian healthy women, aged 75 and over were enrolled into EPIDOS Due to the unavailability of ultrasound devices during the first months of the study and subsequently absent ultrasonic measurements for these participants, N=5898 women were included in this study On treatment at baseline: NR Received treatment during FU: NR
Inclusion criteria	• Caucasian healthy women aged 75 years or over
Exclusion criteria	• undergone a bilateral hip replacement • previously suffered a hip fracture
Bone assessment method	Quantitative ultrasound QUS-heel-BUA; QUS-heel-SOS; QUS-heel-SI (all using Achilles sonometer) Dual X-ray absorptiometry BMD-FN T-score; BMD-total femur T-score
Bone assessment method - baseline	BMD of the proximal femur (femoral neck and total hip BMD) was measured by DXA using the Lunar DPX Plus (GE-Lunar Corp., Madison, WI, USA). QUS of the calcaneus were performed with the Achilles system (GE-Lunar Corp.). The speed of sound (SOS, in m/s) and the ultrasound attenuation over a specific frequency range (BUA, in dB/MHz) were measured by the device, and the and the Stiffness index was calculated based on the combination of both measurements.

Bone assessment method - length of follow up	Mean=3.5 years (representing 21,508 woman-years)
Bone assessment method - follow up	Reference standard: Self-report with fracture confirmed by rheumatologist using preoperative radiographs and surgical reports. A survey of fracture occurrence was conducted every 4 months. Follow-up occurred at 1.5, 2.5 and 3.5 years. Participants were mailed a questionnaire followed by telephone calls (where necessary). If a participant could not be reached, the required information was retrieved from a relative or from her usual doctor. N=538 women (7.2%) were lost to follow-up. *Note: Only data at 3.5 years FU extracted.*
Predicted outcomes	• Hip fracture
Discrimination outcomes	c-statistic/AUC

Study arms

- Hip fracture (N = 227)
- No hip fracture (N = 5671)

Characteristics

Study-level characteristics

Characteristic	Study (N = 5898)
% Female No of events	n = 5898 ; % = 100
Mean age (SD) - Hip fracture Mean (SD)	82.56 (4.53)
Mean age (SD) - Non hip fracture Mean (SD)	80.35 (3.71)
BMI - Hip fracture Mean (SD)	24.64 (3.7)
BMI - Non hip fracture Mean (SD)	25.36 (4.2)

Outcomes

DXA BMD-FN t-score

Outcome	Hip fracture vs No hip fracture, N2 = 5671, N1 = 227
AUC Unadjusted Custom value	0.69 (95%CI 0.66-0.73)

1 AUC - Polarity - Higher values are better

2 AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

3 **DXA BMD-total femur t-score**

Outcome	Hip fracture vs No hip fracture, N2 = 5671, N1 = 227
AUC Unadjusted Custom value	0.64 (95%CI 0.61-0.68)

4 AUC - Polarity - Higher values are better

5 AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

6 **QUS-heel Achillies**

Outcome	Hip fracture vs No hip fracture, N2 = 5671, N1 = 227
AUC - BUA Broadband ultrasound attenuation. Unadjusted Custom value	0.65 (95%CI 0.62-0.69)
AUC - SOS Speed of sound. Unadjusted Custom value	0.65 (95%CI 0.61-0.68)
AUC - SI Stiffness index. Unadjusted Custom value	0.66 (95%CI 0.62-0.7)

7 AUC - Polarity - Higher values are better

1	Achilles sonometer	

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	High *(Study had an overall high risk of bias due to high risk in the patient selection (people with previous hip fracture excluded) and flow and timing (not all participants received the same reference standard as only self-reported fractures were verified by a rheumatologist. Not all participants were included in the analysis (7.2% were lost to follow-up)) domains. All other domains were low risk)*
Overall risk of bias and directness	Directness	Directly applicable *(Participants were recruited from the EPIDOS population (women aged 75 years or older) and subsequently would be at suspected risk of fragility fracture)*

D.1.15 Hillier, 2007

Bibliographic Reference Hillier, Teresa A; Stone, Katie L; Bauer, Doug C; Rizzo, Joanne H; Pedula, Kathryn L; Cauley, Jane A; Ensrud, Kristine E; Hochberg, Marc C; Cummings, Steve R; 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; 2007; vol. 167 (no. 2); 155-60

Study details

Secondary publication of another included study- see primary study for details	NA

Other publications associated with this study included in review	SOF cohort, see Cummings 1994. Participants from SOF cohort also included in Ensrud 2024.
Study type	Prospective cohort study
Study location	USA
Study setting	Community
Study dates	1989 to 1990
Sources of funding	Supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases and by Public Health Service grants from the National Institute on Aging
Study sample	N=4124 older women On treatment at baseline: NR Received treatment during FU: 25%
Inclusion criteria	• community dwelling women • aged 65 years or older • Enrolled in the Study of Osteoporotic Fractures (SOF) cohort
Exclusion criteria	• unable to walk without assistance • with bilateral hip replacements
Bone assessment method	Dual X-ray absorptiometry

	BMD-total hip T-score
Bone assessment method - baseline	BMD of the proximal femur and sub regions (intertrochanter, trochanter, femoral neck, and Ward triangle) were measured using DXA (Hologic QDR 1000; Hologic Inc, Waltham, Mass).
Bone assessment method - length of follow up	Mean=5 years after second BMD measurement (~8-9 years after initial BMD measurement)
Bone assessment method - follow up	Reference standard: Self-report with fracture adjudicated by physician from radiology reports. Participants were contacted every 4 months by postcard, with those who did not respond contacted by telephone, to ascertain incident hip and non-spine fractures. More than 95% of these contacts were completed. Follow-up was completed in the 4,124 women *Note: Data only extracted for initial BMD measurement. Data for non-spine fracture not extracted.*
Predicted outcomes	• Hip fracture
Discrimination outcomes	c-statistic/AUC
Additional comments	Morphometric spine fractures were reported but data was not extracted as this outcome was not relevant to this review

Study arms

- Hip fracture (N = 275)

1 • No hip fracture (N = 3849)

Characteristics

Study-level characteristics

Characteristic	Study (N = 4124)
% Female No of events	n = 4124 ; % = 100
Mean age (SD) Mean (SD)	72 (4)

Outcomes

DXA BMD-TH t-score

Outcome	Hip fracture vs No hip fracture, N2 = 3849, N1 = 275
AUC Unadjusted Custom value	0.73 (95%CI 0.698-0.759)

AUC - Polarity - Higher values are better

Total hip

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group. 95%CIs calculated using equations in Section 3.1, Debray 2018.

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	Moderate *(Study at moderate risk due to high risk in the domain flow and timing. This is because not all participants received the same reference standard as only self-reported fractures were verified by a physician. All participants were included in the analysis and more than 95% of follow-up contacts were completed. All other domains were low risk.)*
Overall risk of bias and directness	Directness	Directly applicable *(Participants were recruited from the SOF population (aged 65 years or older) and subsequently at suspected risk of fragility fracture, however those unable to walk without assistance or with previous bilateral hip replacements were excluded.)*

D.1.16 Krieg, 2006

Bibliographic Reference

Krieg, Marc-Antoine; Cornuz, Jacques; Ruffieux, Christiane; Van Melle, Guy; Buche, Daniel; Dambacher, Maximilian A; Hans, Didier; Hartl, Florian; Hauselmann, Hansjorg J; Kraenzlin, Marius; Lippuner, Kurt; Neff, Maurus; Pancaldi, Pierro; Rizzoli, Rene; Tanzi, Franco; Theiler, Robert; Tyndall, Alan; Wimpfheimer, Claus; Burckhardt, Peter; 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; 2006; vol. 21 (no. 9); 1457-63

Study details

Secondary publication of	NA

another included study- see primary study for details	
Other publications associated with this study included in review	SEMOF cohort, see Popp 2009.
Study type	Prospective cohort study
Study location	Switzerland
Study setting	Community (recruited from official state registries)
Study dates	November 1997 and August 1999
Sources of funding	NR
Study sample	N=7609 elderly ambulatory women ≥70 years of age were included in the SEMOF study On treatment at baseline: 2.8% Received treatment during FU: 6%
Inclusion criteria	• women who were able to walk and independent for daily activities • Enrolled in the Swiss Evaluation of the Methods of Measurement of Osteoporotic Fracture Risk (SEMOF) study
Exclusion criteria	• hip fracture history • bilateral hip replacement

Population subgroups	NA
Bone assessment method	Quantitative ultrasound QUS-heel-BUA; QUS-heel-SOS; QUS-heel-SI (all using Achilles+ sonometer) QUS-heel-BUA; QUS-heel-SOS; QUS-heel-SI (all using Sahara sonometer)
Bone assessment method - baseline	Each participant was assessed by the three QUS devices: Achilles+, Sahara, and DBM Sonic 1200. Achilles+: A heel water-bath ultrasound system that generates a band of frequencies from 200 to 600 kHz. BUA and SOS were measured, and SI was calculated using the following equation: SI= (0.67 × BUA) + (0.28 × SOS) – 420, expressed in percentage of the values obtained by the manufacturer in a young adult population. Sahara: A dry system using an oil-based coupling gel, and the heel is maintained in the box by a specific positioning device. Frequencies of the ultrasounds range from 200 to 600 kHz, and BUA and SOS are measured. QUI is automatically calculated by Sahara from the BUA and the SOS, using the following equation: QUI= 0.41 × (BUA + SOS) − 571, without any unit. DBM Sonic 1200: Measures the propagation of an US pulse through the distal metaphysis of the first phalanges of the last four fingers of the nondominant hand. Amplitude dependent speed of sound (AD-SOS; m/s) is measured and the velocity of the pulse is measured when the amplitude of the US signal reaches a fixed threshold value of 20 mV (trigger level). Before measuring bone tissue, an evaluation of the velocity of the US propagation through the soft tissue is required.
Bone assessment method - length of follow up	Mean=2.9 years (SD 0.8, range 5 days-4.9 years), follow-up of 20,409 women-years.
Bone assessment method - follow up	Reference standard: Self-report with fracture confirmed by medical report. Every 6 months, participants received a questionnaire by mail to register any changes in medical conditions in the intervals, particularly any illness, modification of medication, or fracture, with its precise localization and trauma level. Follow-up was completed in 7602 women (495 (6.5%) did not answer the first follow-up questionnaire and were considered lost,

	and 52 (0.7%) were not measured by the three QUSs and were excluded from the analysis). Self-reported hip fractures were confirmed with a medical report from the physician in charge and were categorized as trochanteric, femoral neck, or limited to the great trochanter. Low-trauma fractures were defined as spontaneous or as the consequence of a fall from standing height or less.
Predicted outcomes	• Hip fracture
Discrimination outcomes	c statistic/AUC
Additional comments	QUS measurements from the phalanges using the DBM Sonic 1200 were deemed not relevant for this review and so this data was not extracted

Study arms

- Hip fracture (N = 80)
- No hip fracture (N = 6982)

Characteristics

Study-level characteristics

Characteristic	Study (N = 7062)
% Female	n = 7062; % = 100

Characteristic	Study (N = 7062)
No of events	
Mean age (SD) Mean (SD)	75.2 (3.1)
BMI (kg/m^2) Mean (SD)	25.9 (4.3)

Outcomes

QUS-heel Sahara

Outcome	Hip fracture vs No hip fracture, N2 = 6982, N1 = 80
AUC - BUA Broadband ultrasound attenuation Custom value	0.72 (95%CI 0.67-0.78)
AUC - SOS Speed of sound Custom value	0.72 (95%CI 0.66-0.78)
AUC - QUI Quantitative ultrasound index Custom value	0.73 (95%CI 0.67-0.79)

1 AUC - Polarity - Higher values are better

2 Sahara sonometer

3 AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

4 **QUS-heel Achilles+**

Outcome	Hip fracture vs No hip fracture, N2 = 6982, N1 = 80
AUC - BUA Broadband ultrasound attenuation Custom value	0.71 (95%CI 0.66-0.77)
AUC - SOS Speed of sound Custom value	0.7 (95%CI 0.64-0.76)
AUC - SI Stiffness index Custom value	0.72 (95%CI 0.66-0.78)

5 AUC - Polarity - Higher values are better

6 Achilles+ sonometer

7 AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

8

9 **Critical appraisal - Critical Appraisal - QUADAS-2**

Overall risk of bias and directness	Risk of Bias	Moderate *(Study had an overall moderate risk of bias due to high risk in the domain flow and timing. This is because*

		not all participants received the same reference standard as only self-reported fractures were verified by medical reports. Not all participants were included in the analysis (6.5% did not answer the first follow-up questionnaire and were considered lost, and 0.7% were not measured by the three QUSs and were excluded from the analysis). The mean follow-up for participants was 2.9 years; however, the minimum follow-up period was 5 days, which would be considered inappropriate. All other domains were low risk.)
Overall risk of bias and directness	Directness	Directly applicable *(Participants were women, aged 70 years or older and subsequently at suspected risk of fragility fracture, however those with hip fracture history or bilateral hip replacement were excluded)*

D.1.17 Kuzma, 2018

Bibliographic Reference Kuzma, Martin; Hans, Didier; Koller, Tomas; Nemethova, Eva; Jackuliak, Peter; Killinger, Zdenko; Resch, Heinrich; Payer, Juraj; 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; 2018; vol. 36 (no. 5); 580-588

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	Prospective cohort study
Study location	Slovakia
Study setting	Secondary (Outpatient clinic)
Study dates	Recruitment between 2009 and January 2013, and follow-up between June through November 2015
Sources of funding	Not reported
Study sample	N=144 eligible post-menopausal women aged≥50 years. 17 were excluded due to outlier BMD/TBS results, leaving N=127 women On treatment at baseline: No Received treatment during FU: 4%
Inclusion criteria	• post-menopausal women aged <50 years • no previous low-impact fracture reported • osteopenic BMD levels • no past or current osteoporosis-specific treatment • no history of premature ovarian failure • a body mass index (BMI) within the range of 15.0–37.5 kg/m2
Exclusion criteria	Not reported
Population subgroups	NA
Bone assessment method	Dual X-ray absorptiometry

	BMD-FN T-score; BMD-LS T-score
Bone assessment method - baseline	BMD at the lumbar spine (LS) and femoral neck (FN) were measured in all participants using LS and FN DXA scans (Hologic©, QDR Discovery). All measurements were performed on the same DXA machine. Phantom scans with repositioning, before and after the hardware change and cross-calibration were performed by the same technologist, whenever the systems software was upgraded until no greater than a 1% difference in mean BMD was observed. TBS was derived from the LS DXA scans and calculated by the analysis software package iNsight® version 1.9.1.0 as part of our clinical routine (Medimaps, Merignac, France).
Bone assessment method - length of follow up	Mean=5.2 years (range 2.7-7.0)
Bone assessment method - follow up	Reference standard: Self-report with fracture confirmed by medical record from surgeon or traumatologist. Low-impact fragility fractures were self-reported via questionnaires and were defined as fractures sustained during a fall from a height no greater than the person's height while standing. The date and localisation site of the fracture was documented for a newly onset low-impact fracture (defined as a fracture sustained during a fall from a height no greater than the person's height while standing), that was confirmed and documented by surgeon or traumatologist, and details on how the fracture was managed were provided. Questionnaires were answered with the assistance of study nurse or physician, and patient reported fractures were documented with the appropriate medical record from a surgeon or a traumatologist. Clinically relevant osteoporotic fractures were included in the analysis. Asymptomatic vertebral fractures were not actively searched for. *Note: Data only extracted for DXA BMD.*
Predicted outcomes	• Major osteoporotic fracture *Note: reported as clinically relevant osteoporotic fractures so may include fractures other than hip, shoulder, clinical vertebral [spine], and forearm*
Discrimination outcomes	c-statistic/AUC

Study arms

- Major osteoporotic fracture (N = 18)
- No major osteoporotic fracture (N = 109)

Characteristics

Study-level characteristics

Characteristic	Study (N = 127)
% Female No of events	n = 127; % = 100
Mean age (SD) - Without fracture Mean (SD)	65.8 (10.6)
Mean age (SD) - With fracture Mean (SD)	68 (9.8)
BMI - Without fracture Mean (SD)	26.1 (4.6)
BMI - With fracture Mean (SD)	25.2 (4.1)

Characteristic	Study (N = 127)
Alcohol intake No of events	n = NA; % = NA
Alcohol intake - Without fracture No of events	n = 0; % = 0
Alcohol intake - With fracture No of events	n = 0; % = 0
Smoking status No of events	n = NA; % = NA
Smoking status - Smoking without fracture No of events	n = 21; % = 19.2
Smoking status - Smoking with fracture No of events	n = 3; % = 16.6
Rheumatoid arthritis or systemic lupus erythematosus - Rheumatoid arthritis without fracture No of events	n = 4; % = 3.6
Rheumatoid arthritis or systemic lupus erythematosus - Rheumatoid arthritis with fracture No of events	n = 0; % = 0

Characteristic	Study (N = 127)
Corticosteroid use - Without fracture No of events	n = 1; % = 0.9
Corticosteroid use - With fracture No of events	n = 0; % = 0

Outcomes

DXA BMD-FN

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 109, N1 = 18
AUC Unadjusted Custom value	0.58 (95%CI 0.45-0.72)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

DXA BMD-LS

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 109, N1 = 18
AUC Unadjusted Custom value	0.39 (95%CI 0.26-0.53)

1 AUC - Polarity - Higher values are better

2 Unadjusted. Lumbar vertebrae scanned not reported.

3 AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

4

5

6 **Critical appraisal - Critical Appraisal - QUADAS-2**

Overall risk of bias and directness	Risk of Bias	High *(Study had an overall high risk of bias due to high risk in the three domains patient selection, reference standard and flow and timing. This is because the exclusion criteria was not reported. It is unclear about use of medical records and whether the reference standard was interpreted without knowledge of the results of the index test. Questionnaires (where fractures were self-reported) were answered with the assistance of study nurse or physician, so this is possible. Knowledge of the index test would be unlikely to introduce bias for the outcome (fracture). Not all participants completed follow-up (7 were excluded due to outliers for BMD and TBS).)*
Overall risk of bias and directness	Directness	HF Directly applicable MOF Partially applicable *(Includes any clinically relevant osteoporotic fracture, no breakdown of types of fracture provided or attempted.)*

7

8 ## D.1.18 Leonhardt, 2020

Bibliographic Reference	Leonhardt, Yannik; May, Pauline; Gordijenko, Olga; Koeppen-Ursic, Veronika A; Brandhorst, Henrike; Zimmer, Claus; Makowski, Marcus R; Baum, Thomas; Kirschke, Jan S; Gersing, Alexandra S; Seifert-Klauss, Vanadin; Schwaiger, Benedikt J;

Opportunistic QCT Bone Mineral Density Measurements Predicting Osteoporotic Fractures: A Use Case in a Prospective Clinical Cohort.; Frontiers in endocrinology; 2020; vol. 11; 586352

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	Prospective cohort study
Study location	Germany
Study setting	Secondary (Hospital)
Study dates	2015 to 2018
Sources of funding	Research funding from the German Research Foundation (Deutsche Forschungsgemeinschaft) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme and the Nvidia Corporation.

Study sample	N=79 eligible participants who had been hospitalized due to low-energy fractures of the spine, proximal femur, proximal humerus, or distal radius. 21 people were excluded due to lack of FU, unreadable CT scans, or altered vertebrae (for example, too many fractured vertebrae), leaving N=58 participants On treatment at baseline: NR Received treatment during FU: NR
Inclusion criteria	• patients who had been hospitalized due to low-energy fractures of the spine, proximal femur, proximal humerus, or distal radius • undergone at least one CT including the lumbar spine either indicated for a suspected acute vertebral fracture or for other reasons during a visit at hospital shortly before a baseline Fracture Liaison Service visit
Exclusion criteria	• insufficient follow-up • heavy impairment of the image quality or alterations in too many vertebrae
Bone assessment method	Quantitative computed tomography vBMD-LS
Bone assessment method - baseline	Participants had scans of the lumbar spine using five different multidetector computed tomography (MDCT) scanners in the hospital (IQon, Brilliance 64 and iCT 256 by Philips Medical Care; Somatom Definition AS+ and Definition AS by Siemens Healthineers), partly with administration intravenous contrast medium (Imeron 400, Bracco). The data was acquired in helical mode with a peak tube voltage of 120 kVp, a slice thickness of 0.9 to 1 mm, and adaptive tube load.
Bone assessment method - length of follow up	≥3 years
Bone assessment method - follow up	Reference standard: New imaging findings or clinical follow up. Follow-up was completed in 58 participants (14 had insufficient follow-up, including lack of CT at the institution and validated external reports, and it was not possible in 7 participants to measure BMD in at least 3 thoracolumbar

	vertebrae at the baseline CT due to due to heavy impairment of the image quality, including metal artefacts or alterations in too many vertebrae including too many fractured vertebrae or vertebrae after kyphoplasty). New low energy fractures were reported either by imaging performed at the institute, or clinical follow-up in the osteoporosis centre of the hospital with patient reporting and/or validated external reports of new vertebral fractures.
Predicted outcomes	• Major osteoporotic fracture (Proximal femur, proximal humerus, spine, or distal radius) Reports all fragility fractures
Discrimination outcomes	c-statistic/AUC

Study arms

- Major osteoporotic fracture (N = 20)
- No major osteoporotic fracture (N = 38)

Characteristics

Study-level characteristics

Characteristic	Study (N = 58)
% Female No of events	n = 42; % = 72
Mean age (SD) Mean (SD)	72.8 (10.69)

Outcomes

QCT vBMD-LS

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 38, N1 = 20
AUC Custom value	0.76 (95%CI 0.61-0.87)

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	High *(Study had an overall high risk of bias due to high risk in the two domains reference standard and flow and timing. This is because there is no specific information on the standard test imaging performed in the institute, nor whether it was conducted by blinded assessors. It is possible that patient reporting might lead to incorrect classifications of incident fractures, however validated external reports were also used as a reference standard and these were less likely to have introduced bias or inaccurate results. Of the 79 eligible participants, 14 (18%) were excluded from analysis due to insufficient follow-up and 7 (9%) due to heavy impairment of image quality. It appears that not all participants received the same reference standard (participants underwent either imaging performed in the institute, or clinical follow-up in the osteoporosis centre of the hospital with patient reporting and/or validated external reports).)*
Overall risk of bias and directness	Directness	HF Directly applicable MOF

		Partially applicable *(Study predicts any fragility fracture so includes more types of fracture than MOF defined as wrist, shoulder, hip, and clinical vertebral [spine] fragility fractures. Breakdown of types of fragility fracture not reported.)*

D.1.19 Leslie, 2014

Bibliographic Reference	Leslie, W D; Aubry-Rozier, B; Lix, L M; Morin, S N; Majumdar, S R; Hans, D; Spine bone texture assessed by trabecular bone score (TBS) predicts osteoporotic fractures in men: the Manitoba Bone Density Program.; Bone; 2014; vol. 67; 10-4

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study included in review	Manitoba Bone Density Program cohort, see also Leslie 2007, Schacter 2017, Zarzour 2024
Study type	Prospective cohort study
Study location	Canada
Study setting	Community

Study dates	Service records were assessed between April 1, 1987 and March 31, 2011
Sources of funding	Speaker fees (paid to facility) from Amgen, Eli Lilly, and Novartis, research grants (paid to facility) from Amgen and Genzyme, and consultant fees from Amgen, Eli Lilly, Novartis, and Merck.
Study sample	N=3620 men aged ≥ 50 years On treatment at baseline: 21.3% Received treatment during FU: NR
Inclusion criteria	• men aged ≥ 50 years • from the Manitoba Bone Density Program associated databases • who had undergone baseline BMD measurement of the spine (L1–L4) and hip (femoral neck) with DXA
Exclusion criteria	Not reported
Bone assessment method	Dual X-ray absorptiometry BMD-FN T-score; BMD-LS T-score; BMD-FN T-score + TBS-LS; BMD-LS + TBS-LS
Bone assessment method - baseline	Participants had BMD measurements recorded for the lumbar spine (L1–L4) and femoral neck using DXA Prodigy scanners (GE Healthcare, Madison, WI, USA) and analysed in accordance with the manufacturer's recommendations (Encore Software 12.4). Three DXA instruments were used and were cross calibrated for BMD using anthropomorphic phantoms with no significant between them. The instruments exhibited stable long-term performance, with a coefficient of variation <0.5%. Anonymized lumbar spine DXA files from the Manitoba Bone Density Program database were used to ensure blinding of investigators, and TBS measurements were performed in the Bone Disease Unit at the Lausanne University Hospital in Lausanne, Switzerland (TBS iNsight Software, Version 1.8, Med-Imaps, Pessac, France). The version of software used had been optimized for Caucasian women.

Bone assessment method - length of follow up	Mean=4.5 years
Bone assessment method - follow up	Reference standard: Diagnostic codes from healthcare records. Follow-up was completed in 3,620 men. In the Province of Manitoba, Canada at each health system contact (for most residents) information is recorded to document the patient's demographics, date and type of service, and diagnostic code(s) using the International Classification of Disease, 9th edition, Clinical Modification (ICD-9-CM) system for physician billing claims. Computerized databases of hospital discharge abstracts coded using ICD-9-CM prior to 2004 and used ICD-10-CA thereafter, and since 1995, a province-wide retail pharmacy database has captured drug dispensations and prescription details. Anonymous linkage across these databases was performed unique scrambled health identification number to create a longitudinal record of health services and outcomes. Health service records were assessed between April 1, 1987, and March 31, 2011, for the presence of fracture codes not associated with severe trauma using validated definitions. These included hip fractures and major osteoporotic fractures (such as hip, clinical vertebral [spine], forearm, and humerus fractures) that occurred after BMD testing.
Predicted outcomes	• Major osteoporotic fracture (Hip, shoulder, clinical vertebral [spine], forearm) • Hip fracture
Discrimination outcomes	c-statistic/AUC

Study arms

- Major osteoporotic fracture (N = 183)
- No major osteoporotic fracture (N = 3437)
- Hip fracture (N = 46)

1 • No hip fracture (N = 3574)

2

3 **Characteristics**

4 **Study-level characteristics**

Characteristic	Study (N = 3620)
% Female No of events	n = 0; % = 0
Mean age (SD) Mean (SD)	67.6 (9.8)
BMI (kg/m^2) Mean (SD)	27.1 (4.4)
Alcohol intake - Alcohol or substance abuse No of events	n = 113; % = 3.1
Previous fracture No of events	n = 592; % = 16.4
Chronic obstructive airways disease or asthma - Chronic obstructive pulmonary disease No of events	n = 474; % = 13.1
Rheumatoid arthritis or systemic lupus erythematosus - Rheumatoid arthritis	n = 191; % = 5.3

Characteristic	Study (N = 3620)
No of events	
Corticosteroid use - Recent glucocorticoid use No of events	n = 688; % = 19

Outcomes

DXA BMD-FN t-score

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 3437, N1 = 183	Hip fracture vs No hip fracture, N2 = 3574, N1 = 46
AUC Custom value	0.68 (95%CI 0.64-0.72)	0.77 (95%CI 0.70-0.84)

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

DXA BMD-LS t-score

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 3437, N1 = 183	Hip fracture vs No hip fracture, N2 = 3574, N1 = 46
AUC Custom value	0.64 (95%CI 0.59-0.68)	0.68 (95%CI 0.60-0.75)

1 AUC - Polarity - Higher values are better

2 L1-L4

3 AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

4 **DXA BMD-FN t-score + Trabecular Bone Score-LS**

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 3437, N1 = 183	Hip fracture vs No hip fracture, N2 = 3574, N1 = 46
AUC Custom value	0.68 (95%CI 0.64-0.72)	0.78 (0.71-0.85)

5 AUC - Polarity - Higher values are better

6 L1-L4

7 DXA of BMD at femoral neck and trabecular bone score at the lumbar spine

8 AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

9 **DXA BMD-LS t-score + Trabecular Bone Score-LS**

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 3437, N1 = 183	Hip fracture vs No hip fracture, N2 = 3574, N1 = 46
AUC Custom value	0.64 (95%CI 0.60-0.68)	0.70 (95%CI 0.62-0.78)

10 AUC - Polarity - Higher values are better

11 L1-L4

1	DXA of BMD at lumbar spine and trabecular bone score at the lumbar spine	

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	High *(Study had an overall high risk of bias due to high risk in the two domains patient selection and reference standard. This is because it is unclear whether a random sample or consecutive participants were enrolled, and the exclusion criteria was not reported. Fracture occurrence was identified using diagnostic codes from healthcare records, and so is an indirect way of ascertaining fracture occurrence.)*
Overall risk of bias and directness	Directness	Directly applicable

D.1.20 Leslie, 2007

Bibliographic Reference	Leslie, William D; Tsang, James F; Caetano, Patricia A; Lix, Lisa M; Effectiveness of bone density measurement for predicting osteoporotic fractures in clinical practice.; The Journal of clinical endocrinology and metabolism; 2007; vol. 92 (no. 1); 77-81

Study details

Secondary publication of another included study- see primary study for details	NA

Other publications associated with this study included in review	Manitoba Bone Density Program cohort, see also Leslie 2014, Schacter 2017, Zarzour 2024
Study type	Prospective cohort study
Study location	Canada
Study setting	Community
Study dates	Bone densitometry was performed between May 1998 and October 31, 2002, and the observation period ended March 31, 2004.
Sources of funding	This work was supported in part by an unrestricted educational grant from the CHAR/GE Healthcare Development Awards Programme.
Study sample	N=16,505 women aged ≥50 years On treatment at baseline: NR Received treatment during FU: NR
Inclusion criteria	• women aged ≥50 years • who had baseline lumbar spine and proximal femur bone densitometry performed before October 31, 2002, with one of the program's primary instruments • who had medical coverage with Manitoba Health during the observation period ending March 31, 2004
Exclusion criteria	• test results analysed with earlier software versions (before May 1998) that did not provide total hip measurements

Bone assessment method	Dual X-ray absorptiometry BMD-FN T-score; BMD-LS T-score; BMD-total hip T-score
Bone assessment method - baseline	DXA measurements were performed with a pencil-beam instrument (Lunar DPX; GE Lunar) before the year 2000, and after this date fan-beam instruments were used (Lunar Prodigy; GE Lunar). Densitometers underwent daily assessment of stability using an anthropomorphic spine phantom, and each showed stable long-term performance (coefficient of variation=sd/mean<0.5%). In vivo precision was as follows; coefficient of variation 1.7% for L1– 4 from 198 spine scan pairs and 1.1% for the total hip from 193 hip scan pairs)
Bone assessment method - length of follow up	Mean (SD) observation period=3.2 (1.5) years
Bone assessment method - follow up	Reference standard: Diagnostic codes from healthcare records. Follow-up was completed in the 16,505 women. A longitudinal record of health services and outcomes was created by linking the Manitoba Health computerized databases of physician billing claims and hospital discharges for all residents of the province eligible to receive health services, to the BMD database. Each health system contact includes information on a patient's demographics, date and type of service, and diagnoses, which are coded using the International Classification of Disorders-9-Clinical Modification (ICD-9-CM). This allowed for the identification of the presence of fracture codes including osteoporotic fractures. Hip fractures and wrist fractures also had to be accompanied by a site-specific fracture reduction, fixation, or casting code so as to exclude less severe fractures such as isolated trochanteric fractures not requiring surgical fixation and distal radius fractures not requiring immobilization. Specific fracture sites of interest were the hip (ICD-9-CM 820 – 821), spine (ICD-9-CM 805), wrist (ICD-9-CM 813), and proximal humerus (ICD-9-CM 812). Hip, spine, wrist, and proximal humerus were collectively designated as osteoporotic fractures. Fractures associated with nonaccidental ICD-9-CM trauma codes (ICD-9-CM E800-E879 and E890-E99) were excluded.
Predicted outcomes	• Major osteoporotic fracture (hip, shoulder, clinical vertebral [spine], forearm) • Hip fracture

Discrimination outcomes	c-statistic/AUC

Study arms

- Major osteoporotic fracture (N = 765)
- No major osteoporotic fracture (N = 15740)
- Hip fracture (N = 189)
- No hip fracture (N = 16316)

Characteristics

Study-level characteristics

Characteristic	Study (N = 16505)
% Female No of events	n = 16505 ; % = 100
Mean age (SD) Mean (SD)	65 (9)
Ethnicity - white	n = 16210 ; % = 98.2

Characteristic	Study (N = 16505)
No of events	
Ethnicity - Asian No of events	n = 217 ; % = 1.3
Ethnicity - Hispanic No of events	n = 4 ; % = 0.02
Ethnicity - Black No of events	n = 47 ; % = 0.3
Ethnicity - Other No of events	n = 27 ; % = 0.2

Outcomes

DXA BMD-FN t-score

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 15740, N1 = 765	Hip fracture vs No hip fracture, N2 = 16316, N1 = 189
AUC Custom value	0.69 (95%CI 0.68-0.71)	0.79 (95%CI 0.76-0.83)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group.

DXA BMD-LS t-score

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 15740, N1 = 765	Hip fracture vs No hip fracture, N2 = 16316, N1 = 189
AUC Custom value	0.65 (95%CI 0.63-0.67)	0.66 (95%CI 0.62-0.70)

AUC - Polarity - Higher values are better

L1-L4

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group.

DXA BMD-TH t-score

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 15740, N1 = 765	Hip fracture vs No hip fracture, N2 = 16316, N1 = 189
AUC Custom value	0.71 (95%CI 0.69-0.73)	0.82 (95%CI 0.79-0.85)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group.

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	Moderate *(Study had an overall moderate risk of bias due to high risk in the domain reference standard. Fracture ascertainment used diagnostic codes from healthcare records, which is an indirect way of confirming fractures. All other domains were low risk.)*
Overall risk of bias and directness	Directness	Directly applicable *(Participants were women aged ≥ 50 years. No further details were provided; however, it is likely participants were postmenopausal and subsequently at suspected risk of fragility fracture (mean age of 65 years).)*

D.1.21 Popp, 2009

Bibliographic Reference Popp, A W; Senn, C; Franta, O; Krieg, M A; Perrelet, R; Lippuner, K; 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; 2009; vol. 20 (no. 8); 1393-9

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study	SEMOF cohort, see also Krieg 2006.

included in review	
Study type	Prospective cohort study
Study location	Switzerland
Study setting	Community (recruitment from official state registries)
Study dates	Recruitment between January 1998 and April 2000, and follow-up completed in October 2002
Sources of funding	Not reported
Study sample	N=701 eligible elderly women aged 70 to 80 years. 64 women were excluded due to either no baseline DXA measurement (n=14) or lost to FU (n=50). N=637 women analysed. On treatment at baseline: 5% fracture patients vs 12% no fracture patients Received treatment during FU: NR
Inclusion criteria	• elderly women aged 70 to 80 years • able to walk • independent for their daily activities
Exclusion criteria	• women with a history of hip fracture or bilateral hip replacement
Bone assessment method	Dual X-ray absorptiometry BMD-FN T-score, adjusted for age and fracture history; BMD-LS ICSD T-score, adjusted for age and fracture history; BMD-total hip T-score, adjusted for age and fracture history

Bone assessment method - baseline	Participants had BMD measured at the lumbar spine (LS, first to fourth lumbar vertebrae) and at the non-dominant (non-fractured) total hip, femoral neck, trochanter, distal tibial diaphysis (T-DIA), and distal tibial epiphysis (T-EPI) using DXA (Hologic QDR 4500A™, Hologic, Bedford, MA, USA). For repeated longitudinal measurements, mean precision error of this method was 1.4% and 2.1% for T-EPI and T-DIA, respectively. The region of interest was the area of 120 mm height and 129 mm width, starting 10 mm above the top of the ankle joint space. T-EPI corresponded to the distal 40 mm of the ROI and T-DIA to the proximal 40 mm of the region of interest. For repeated longitudinal measurements, mean precision error of this method was 1.4% for T-EPI, and 2.1% for T-DIA.
Bone assessment method - length of follow up	Mean=2.8 years (SD 0.6) years, range 0.12 to 3.94, corresponding to1,786 women-years of follow-up.
Bone assessment method - follow up	Reference standard: Self-report with fracture confirmed by questionnaire to family practitioner or hospital in charge. Participants returned a questionnaire every 6 months by mail, registering any health changes, including any fracture that occurred during the time interval between two questionnaires. For every fracture, the exact localisation and trauma intensity were to be indicated. Low trauma or fragility fractures were defined as either spontaneous or consecutive to a fall from standing height or less. Follow-up was completed in 637 participants (14 had no baseline DXA measurement and 50 (7.1%) were lost to follow-up).
Predicted outcomes	• Major osteoporotic fracture *Note: Reported as fragility fracture, includes ankle, elbow, clavicle, rib, patella, and tibia fracture in addition to forearm, spine, proximal humerus, hip/pelvis fracture.*
Discrimination outcomes	c-statistic/AUC

Study arms

- Major osteoporotic fracture in women (N = 58)

- No major osteoporotic fracture in women (N = 579)

Characteristics

Study-level characteristics

Characteristic	Study (N = 637)
% Female No of events	n = 637; % = 100
Mean age (SD) Mean (SD)	76 (3)
BMI (kg/m^2) Mean (SD)	25.8 (4.3)
Previous fracture No of events	n = 331; % = 52

Arm-level characteristics

Characteristic	Major osteoporotic fracture in women (N = 58)	No major osteoporotic fracture in women (N = 579)
Hormone replacement therapy Current oestrogen use Sample size	n = 3; % = 5	n = 69; % = 12

Outcomes

DXA BMD-FN t-score

Outcome	Major osteoporotic fracture in women vs No major osteoporotic fracture in women, N2 = 579, N1 = 58
AUC Custom value	0.65 (95%CI 0.58-0.72)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

DXA BMD-LS t-score

Outcome	Major osteoporotic fracture in women vs No major osteoporotic fracture in women, N2 = 579, N1 = 58
AUC - L1-L4 Custom value	0.63 (95%CI 0.56-0.70)
AUC - ISCD T-score Custom value	0.65 (95%CI 0.58-0.72)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

DXA BMD-trochanter t-score

Outcome	Major osteoporotic fracture in women vs No major osteoporotic fracture in women, N2 = 579, N1 = 58
AUC Custom value	0.64 (95%CI 0.57-0.72)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

DXA BMD-TH t-score

Outcome	Major osteoporotic fracture in women vs No major osteoporotic fracture in women, N2 = 579, N1 = 58
AUC Custom value	0.64 (95%CI 0.56-0.71)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	High *(High ROB due to concerns in reference standard and flow and timing domain. Self-reported MOF occurrence confirmed only by questionnaire to family practitioner or relevant hospital. Not all participants received same reference standard as only reported fractures were confirmed by questionnaire.)*
Overall risk of bias and directness	Directness	HF Directly applicable

		MOF Partially applicable *(Includes all fragility fractures, ~10% of which were ankle, elbow, clavicle, rib, patella, or tibia fractures.)*

D.1.22 Schacter, 2017

Bibliographic Reference Schacter, G I; Leslie, W D; Majumdar, S R; Morin, S N; Lix, L M; Hans, D; 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; 2017; vol. 28 (no. 11); 3199-3203

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study included in review	Manitoba Bone Density Program cohort, see also Leslie 2007/2014, and Zarzour 2024
Study type	Prospective cohort study
Study location	Canada

Study setting	Community
Study dates	1999 to 2011
Sources of funding	No external funding
Study sample	N=52,804 (47,736 women and 4348 men) aged ≥40 years On treatment at baseline: NR Received treatment during FU: NR
Inclusion criteria	• women and men aged ≥40 years • from Manitoba, Canada • with available baseline spine DXA data from the Manitoba Bone Density Program
Exclusion criteria	Not reported
Bone assessment method	Dual X-ray absorptiometry BMD-LS
Bone assessment method - baseline	Participants had DXA of the spine acquired using a single narrow fan-beam scanner configuration (Prodigy, GE Healthcare, Madison, WI, USA). All three instruments used for this study exhibited stable long-term performance (coefficient of variation [CV] < 0.5%) and satisfactory in vivo precision. Short-term reproducibility (CV) for TBS was 2.1% and for spine BMD was 1.7% in 92 individuals with repeat spine DXA scans performed within 28 days.
Bone assessment method - length of follow up	Mean=5.0 years (men) Mean=6.0 years (women)

Bone assessment method - follow up	Reference standard: Diagnostic codes from healthcare records. Follow-up was completed in 52,084 participants. Almost all residents in the Province of Manitoba, Canada, are provided health services through a single public health care system which includes demographics, date and type of service, and diagnoses from physician billing claims (inpatient, outpatient, and private office) and hospital discharge abstracts. These are coded using the International Classification of Disease 9th and 10th edition, Clinical Modification (ICD-9-CM and ICD-10-CM) systems. Subsequently the creation of a longitudinal record of health services and outcomes was possible through anonymous linkage of these databases to the BMD database using a unique scrambled health identification number, and the linkage to health service records was used to identify incident MOFs and hip fractures. Hip and forearm fractures had to be accompanied by a site-specific fracture reduction, fixation, or casting code, which enhances the diagnostic and temporal specificity of an acute fracture.
Predicted outcomes	• Major osteoporotic fracture for men • Major osteoporotic fracture for women • Hip fracture for men • Hip fracture for women
Discrimination outcomes	c-statistic/AUC

Study arms

- Major osteoporotic fracture in men (N = 214)
- No major osteoporotic fracture in men (N = 4134)
- Major osteoporotic fracture in women (N = 2895)
- No major osteoporotic fracture in women (N = 44841)

- Hip fracture in men (N = 47)
- No hip fracture in men (N = 4301)
- Hip fracture in women (N = 694)
- No hip fracture in women (N = 47042)

Characteristics

Study-level characteristics

Characteristic	Study (N = 52084)
% Female No of events	n = 47736; % = 91.6
Mean age (SD) - Men Mean (SD)	64 (12)
Mean age (SD) - Women Mean (SD)	63 (11)
BMI - Men Mean (SD)	26.8 (5.2)
BMI - women Mean (SD)	27.1 (4.5)

Outcomes

DXA BMD-LS t-score - Major osteoporotic fracture

Outcome	Major osteoporotic fracture in men vs No major osteoporotic fracture in men, N2 = 4134, N1 = 214	Major osteoporotic fracture in women vs No major osteoporotic fracture in women, N2 = 44841, N1 = 2895
AUC Custom value	0.637 (95%CI 0.601-0.672)	0.662 (95%CI 0.651-0.672)

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

DXA BMD-LS t-score - Hip fracture

Outcome	**Hip fracture in men vs No hip fracture in men, N2 = 4301, N1 = 47**	**Hip fracture in women vs No hip fracture in women, N2 = 47042, N1 = 694**
AUC Custom value	0.678 (95%CI 0.602-0.754)	0.677 (95%CI 0.656-0.698)

AUC - Polarity - Higher values are better

Critical appraisal - Critical Appraisal - QUADAS-2

Overall risk of bias and directness	Risk of Bias	Moderate *(Study had an overall moderate risk of bias due to unclear patient selection in the population (no info about recruitment/exclusions) and high risk in the domain reference standard. This is because anonymous linkage of databases to the BMD database and health service records was used to identify incident MOFs and hip*

		fractures, and the committee agreed that databases were not a reliable reference standard. All other domains were low risk.)
Overall risk of bias and directness	Directness	Directly applicable

D.1.23 Zarzour, 2024

Bibliographic Reference Zarzour, Fatima; Leslie, William D; 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; 2024; vol. 27 (no. 3); 101502

Study details

Secondary publication of another included study- see primary study for details	NA
Other publications associated with this study included in review	Manitoba Bone Density Program cohort, see also Leslie 2007/2014, and Schacter 2017
Study type	Prospective cohort study
Study location	Canada

Study setting	Community
Study dates	1997 to 2018
Sources of funding	No external funding body
Study sample	N=39,727 (3571 men/36,156 women) adults aged ≥40 years On treatment at baseline: NR Received treatment during FU: NR
Inclusion criteria	• individuals aged ≥40 years • with baseline spine and hip DXA assessment • who were registered for healthcare in Manitoba
Exclusion criteria	• evidence of structural artifact in the lumbar spine BMD (defined as adjacent vertebral T-score differences>1)
Bone assessment method	Dual X-ray absorptiometry BMD-FN T-score; BMD-LS T-score; BMD-TH T-score
Bone assessment method - baseline	Participants had spine and hip DXA scans performed with a cross-calibrated narrow fan-beam DXA instruments (Prodigy before November 2012, iDXA from November 2012 onwards, GE Healthcare, Madison, WI, USA). Scans were analysed in accordance with manufacturer recommendations, and quality assurance was monitored by a medical physicist. Stable machine performance over time was confirmed, with a coefficient of variation <0.5 %). All reporting physicians and DXA charge technologists were required to be ISCD-certified.
Bone assessment method - length of follow up	Mean=8.7 years (SD 5.2)

Bone assessment method - follow up	Reference standard: Diagnostic codes from healthcare records. Follow-up was completed in 39,727 participants. Hospital diagnoses and procedures were coded using the International Classification of Diseases (ICD), Ninth Revision, Clinical Modification, or Tenth Revision, Canadian Enhancements (ICD-10-CA). The diagnostic codes of fractures from Manitoba Health records, and physician billing claims that were coded using ICD Ninth Revision, Clinical Modification were evaluated. Hip and forearm fractures had to be linked with site-specific fracture reduction, fixation, or casting codes to qualify as acute fracture events, and to avoid potential misclassification there should be no documented same-site fracture in the six months preceding an incident fracture diagnosis. Incidents of major osteoporotic fractures were included encompassing hip, clinical vertebral [spine], forearm, and humerus. Fractures associated with high-trauma codes were excluded.
Predicted outcomes	• Major osteoporotic fracture (Hip, shoulder, clinical vertebral [spine], forearm) • Hip fracture
Discrimination outcomes	c-statistic/AUC

Study arms

Major osteoporotic fracture (N = 4809)

No major osteoporotic fracture (N = 34918)

Hip fracture (N = 1085)

1 No hip fracture (N = 38642)

2

3 **Characteristics**

4 **Study-level characteristics**

Characteristic	Study (N = 39727)
% Female No of events	n = 36156; % = 91
Mean age (SD) Mean (SD)	62.7 (10.5)

5

6 **Outcomes**

7 **DXA BMD-FN t-score**

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 36127, N1 = 3600	Hip fracture vs No hip fracture, N2 = 38642, N1 = 1085
AUC Custom value	0.66 (95%CI 0.651-0.670)	0.765 (95%CI 0.752-0.779)

8 AUC - Polarity - Higher values are better

9 AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

10 **DXA BMD-LS t-score**

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 36127, N1 = 3600	Hip fracture vs No hip fracture, N2 = 38642, N1 = 1085
AUC Custom value	0.645 (95%CI 0.636-0.654)	0.668 (95%CI 0.651-0.684)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

DXA BMD-TH t-score

Outcome	Major osteoporotic fracture vs No major osteoporotic fracture, N2 = 36127, N1 = 3600	Hip fracture vs No hip fracture, N2 = 38642, N1 = 1085
AUC Custom value	0.666 (95%CI 0.656-0.675)	0.765 (95%CI 0.751-0.779)

AUC - Polarity - Higher values are better

AUC discriminates between people who will sustain a fracture and those who will not. n1=fractured group, n2=non-fractured group

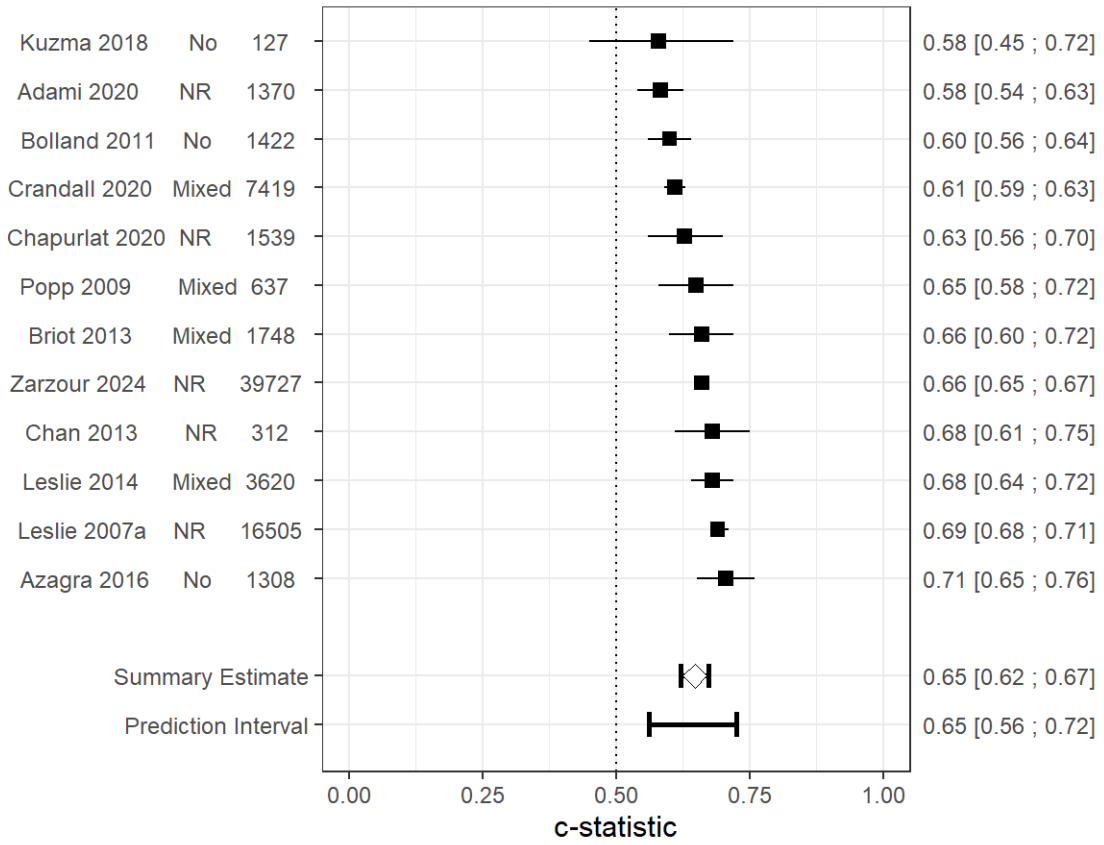
Critical appraisal - Critical Appraisal - QUADAS-2

Section	Question	Answer
Overall risk of bias and directness	Risk of Bias	High *(Study had an overall high risk of bias due to high risk in both reference standard (diagnostic codes of fractures from Manitoba Health records, and physician billing claims were evaluated and linked with site-specific fracture reduction, fixation, or casting codes to qualify as acute fracture events) and flow and timing domains (not all patients received same reference standard)*
Overall risk of bias and directness	Directness	Directly applicable

1

Appendix E Forest plots for accuracy of bone assessment methods for predicting fragility fracture

Figure 2: Forest plot for meta-analysis of DXA of femoral neck to predict major osteoporotic fracture

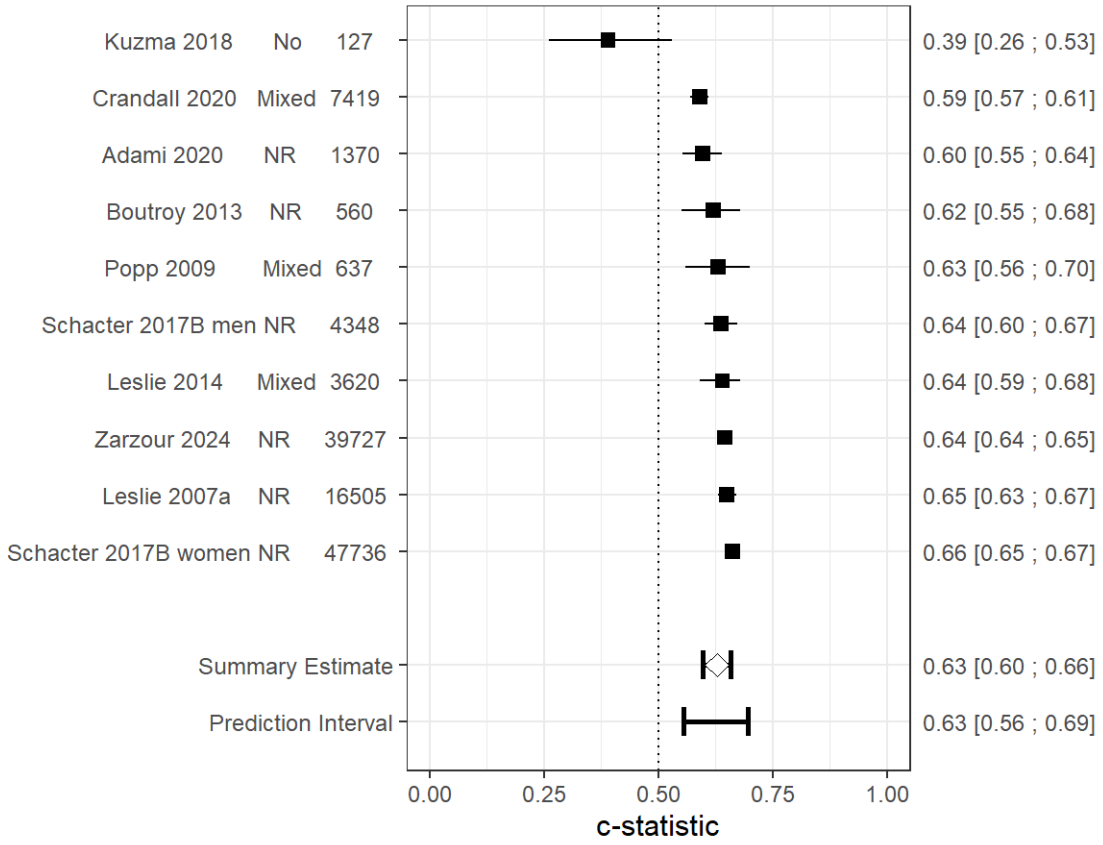


Study, treatment history at baseline, number of participants.

Mixed, some participants received anti-osteoporosis treatment at baseline; No, participants did not receive anti-osteoporosis treatment at baseline; NR, study does not report use of anti-osteoporosis treatment at baseline.

Frequentist model, heterogeneity statistics: I^2=85.3%; τ^2=0.024

Figure 3: Forest plot for meta-analysis of DXA of lumbar spine to predict major osteoporotic fracture

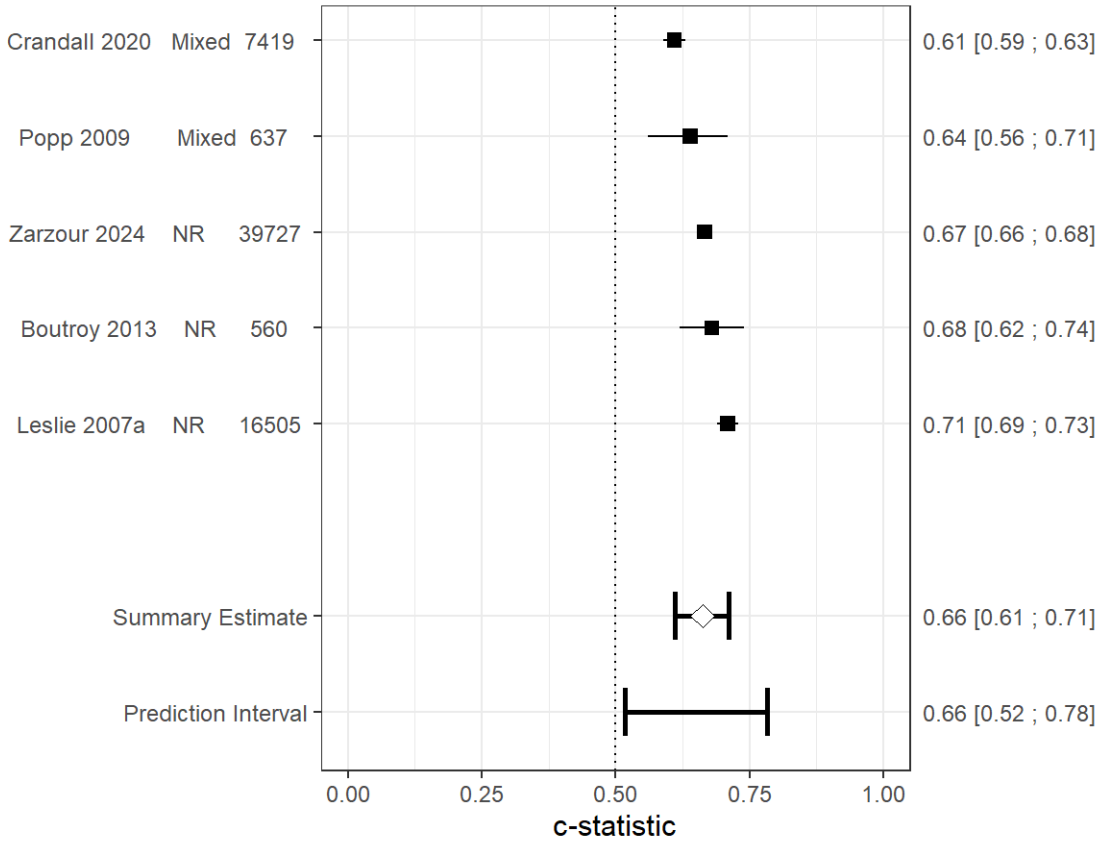


Study, treatment history at baseline, number of participants.

Mixed, some participants received anti-osteoporosis treatment at baseline; No, participants did not receive anti-osteoporosis treatment at baseline; NR, study does not report use of anti-osteoporosis treatment at baseline.

Frequentist model, heterogeneity statistics: I^2=86.01%; τ^2=0.014.

Figure 4: Forest plot for meta-analysis of DXA of total hip to predict major osteoporotic fracture

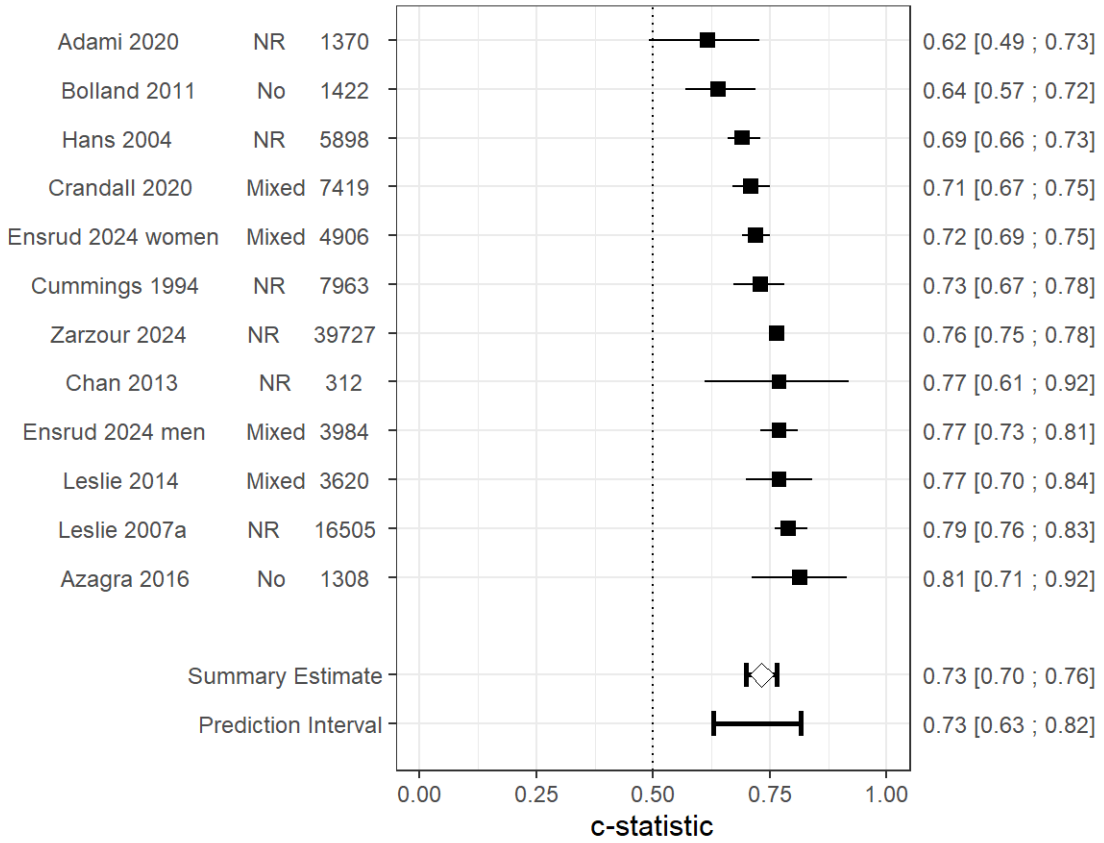


Study, treatment history at baseline, number of participants.

Mixed, some participants received anti-osteoporosis treatment at baseline; No, participants did not receive anti-osteoporosis treatment at baseline; NR, study does not report use of anti-osteoporosis treatment at baseline.

Frequentist model, heterogeneity statistics: I^2=92.1%; τ^2=0.030.

Figure 5: Forest plot for meta-analysis of DXA of femoral neck to predict hip fracture

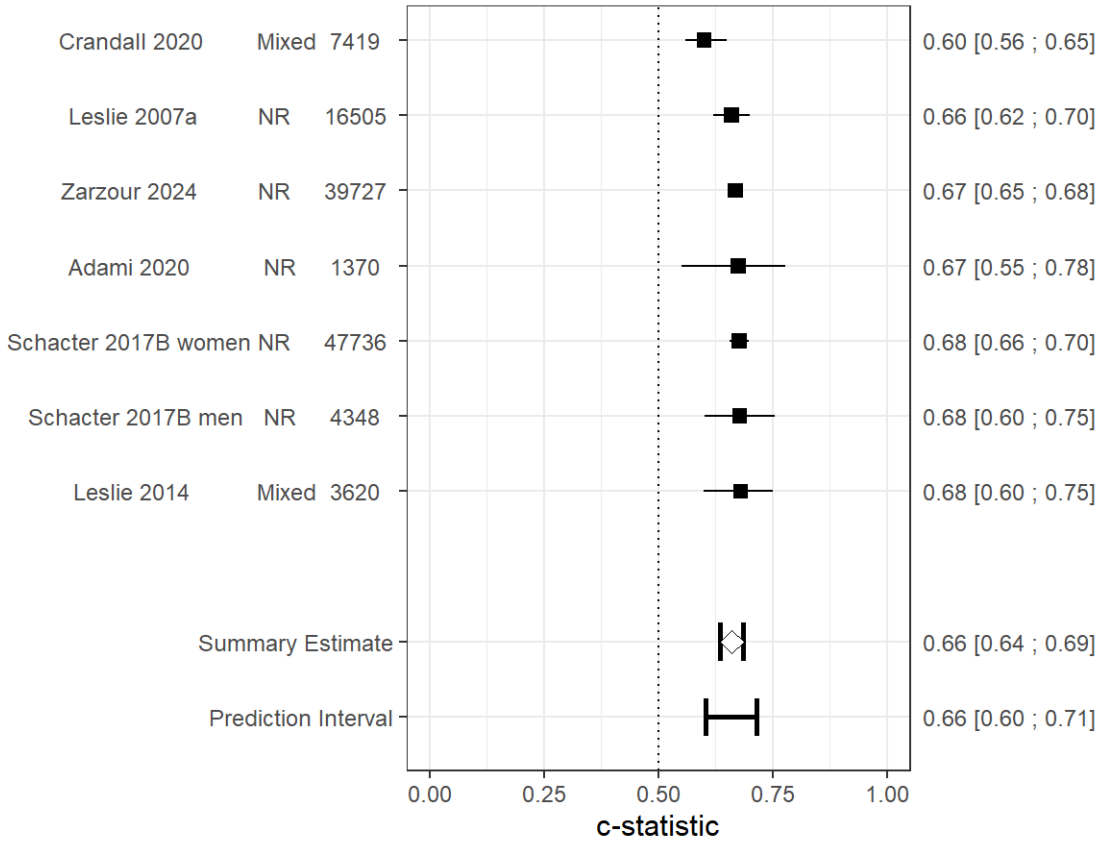


Study, treatment history at baseline, number of participants.

Mixed, some participants received anti-osteoporosis treatment at baseline; No, participants did not receive anti-osteoporosis treatment at baseline; NR, study does not report use of anti-osteoporosis treatment at baseline.

Frequentist model, heterogeneity statistics: I^2=78.9%; τ^2=0.041.

Figure 6: Forest plot for meta-analysis of DXA of lumbar spine to predict hip fracture

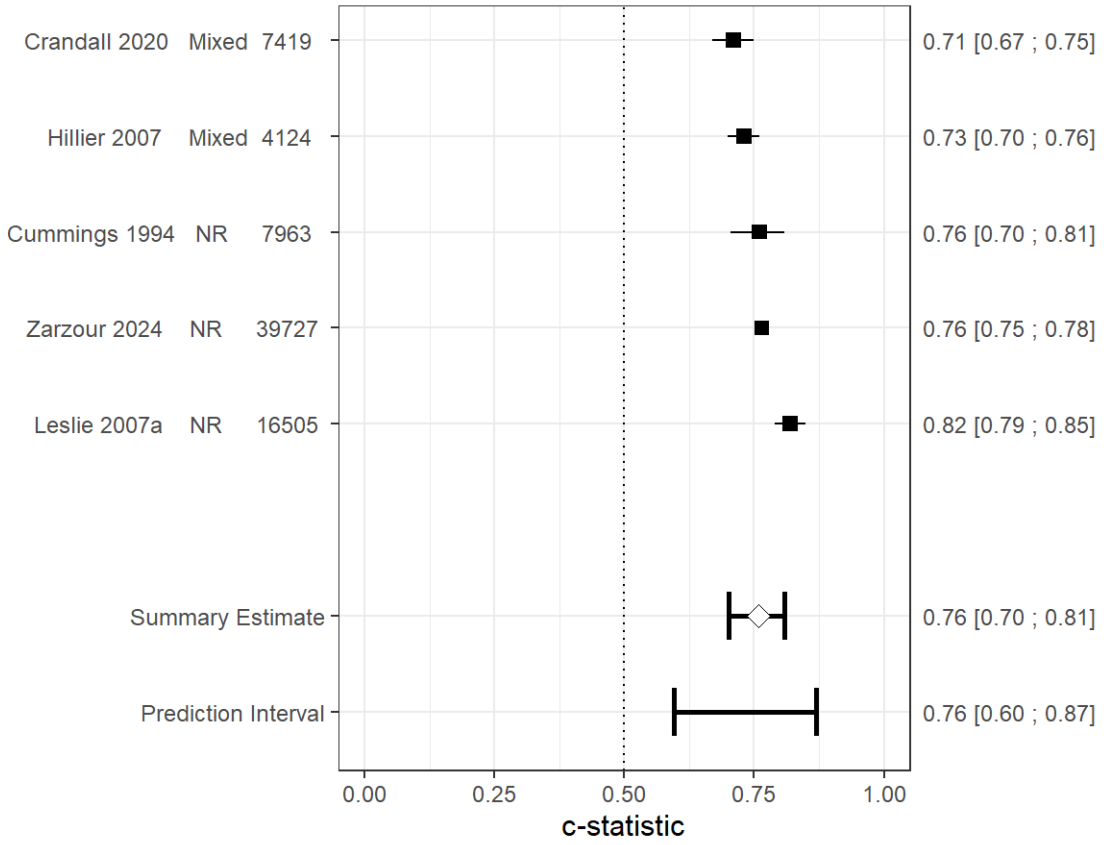


Study, treatment history at baseline, number of participants.

Mixed, some participants received anti-osteoporosis treatment at baseline; No, participants did not receive anti-osteoporosis treatment at baseline; NR, study does not report use of anti-osteoporosis treatment at baseline.

Frequentist model, heterogeneity statistics: I^2=53.7%; τ^2=0.008.

Figure 7: Forest plot for meta-analysis of DXA of total hip to predict hip fracture

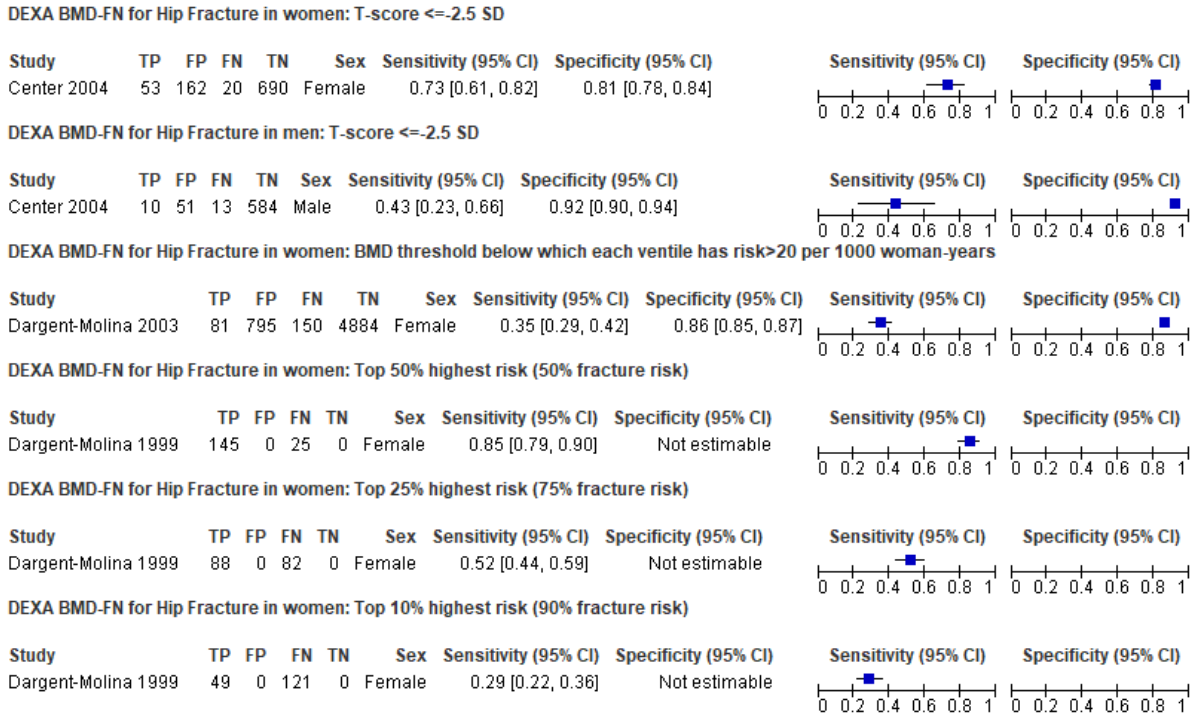


Study, treatment history at baseline, number of participants.

Mixed, some participants received anti-osteoporosis treatment at baseline; No, participants did not receive anti-osteoporosis treatment at baseline; NR, study does not report use of anti-osteoporosis treatment at baseline.

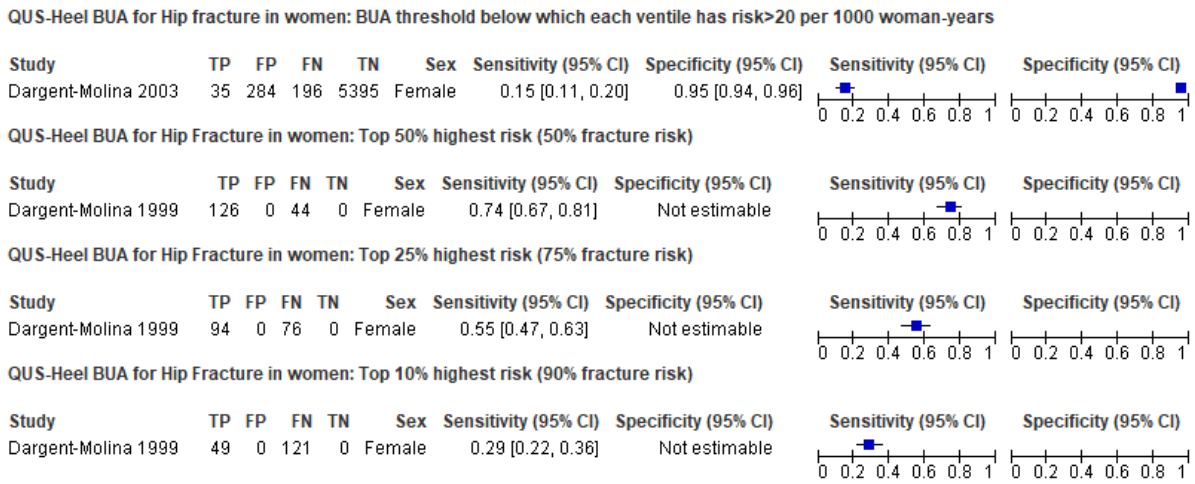
Frequentist model, heterogeneity statistics: I^2=87.2%; τ^2=0.045.

Figure 8: Forest plots for discriminatory power of DXA of femoral neck at various fracture risk thresholds to predict hip fracture



Note: Dargent-Molina 1999 adjusted BMD for age. Raw data for Dargent-Molina 1999 and 2003 estimated using RevMan 5.4.1 calculator from reported sensitivity/specificity. None of the studies reported whether the participants were on antiosteoporosis treatment before or during study.

Figure 9: Forest plots for discriminatory power of QUS of heel at various fracture risk thresholds to predict hip fracture



Note: Dargent-Molina 1999 adjusted BMD for age. Raw data for Dargent-Molina 1999 and 2003 estimated using RevMan 5.4.1 calculator from reported sensitivity/specificity. None of the studies reported whether the participants were on antiosteoporosis treatment before or during study.

Appendix F QUADAS-2 tables for accuracy of bone assessment methods for predicting fragility fracture

Table 9: QUADAS-2 tables for bone assessment methods to predict major osteoporotic fracture

Study	RISK OF BIAS				APPLICABILITY		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Adami 2020	High	Low	High	Low	Low	Low	High
Azagra 2016	Low	Low	High	High	Low	Low	Low
Bolland 2011	Low	Low	High	High	Low	Low	Low
Boutroy 2013	Low	Low	High	High	Low	Low	Low
Briot 2013	Low	Low	Low	High	Low	Low	Low
Chan 2013	High	Low	Unclear	High	Low	Low	High
Chapurlat 2020	High	Low	High	High	Low	Low	Low
Crandall 2020	High	Low	High	High	Low	Low	Low
Kuzma 2018	High	Low	High	High	Low	Low	High
Leonhardt 2020	Low	Low	High	High	Low	Low	High
Leslie 2007	Low	Low	High	Low	Low	Low	Low
Leslie 2014	High	Low	High	Low	Low	Low	Low
Popp 2009	Low	Low	High	High	Low	Low	High
Schacter 2017B	Unclear	Low	High	Low	Low	Low	Low
Zarzour 2024	Low	Low	High	High	Low	Low	Low

Figure 10: QUADAS-2 domains for major osteoporotic fracture

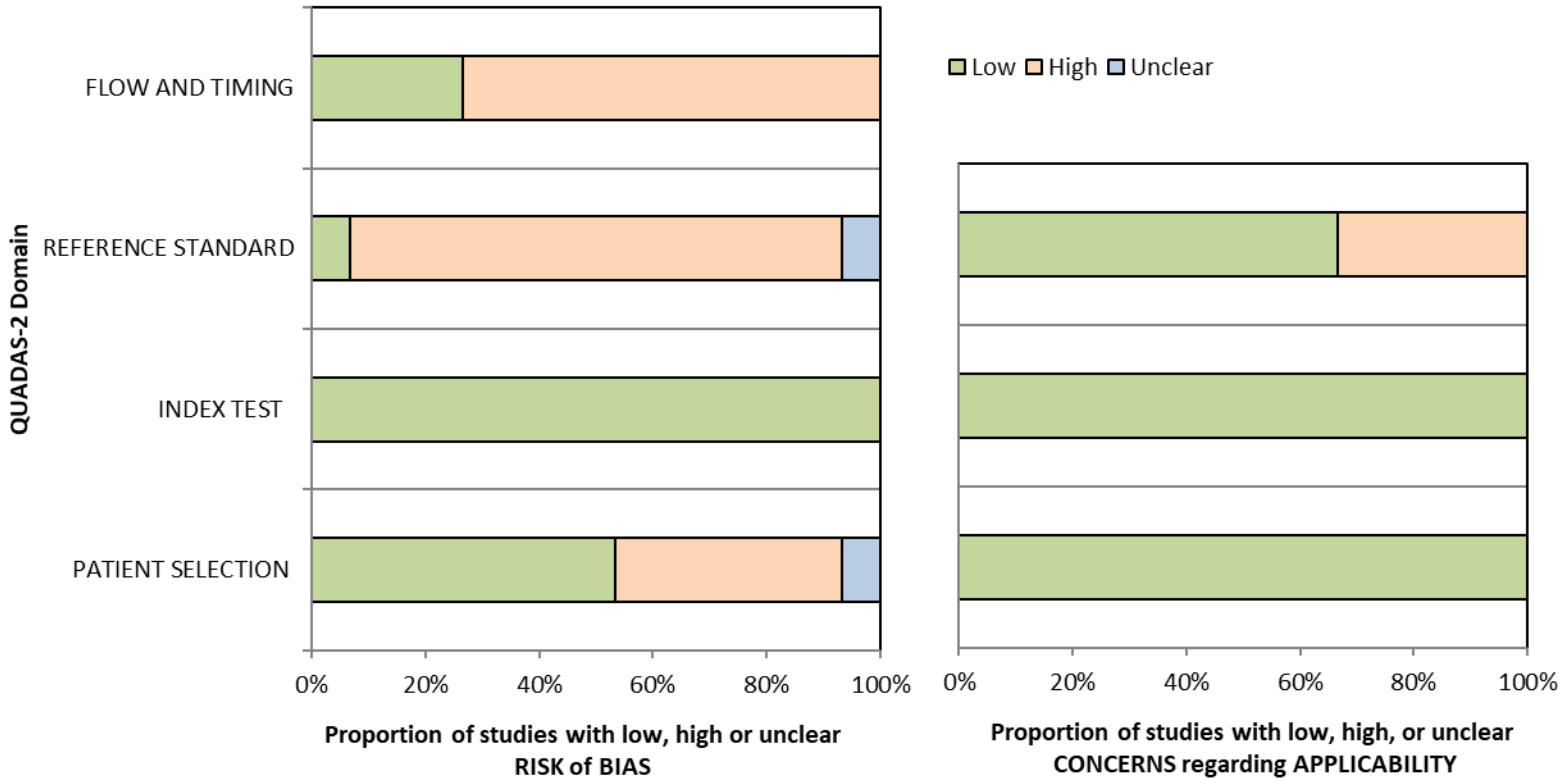
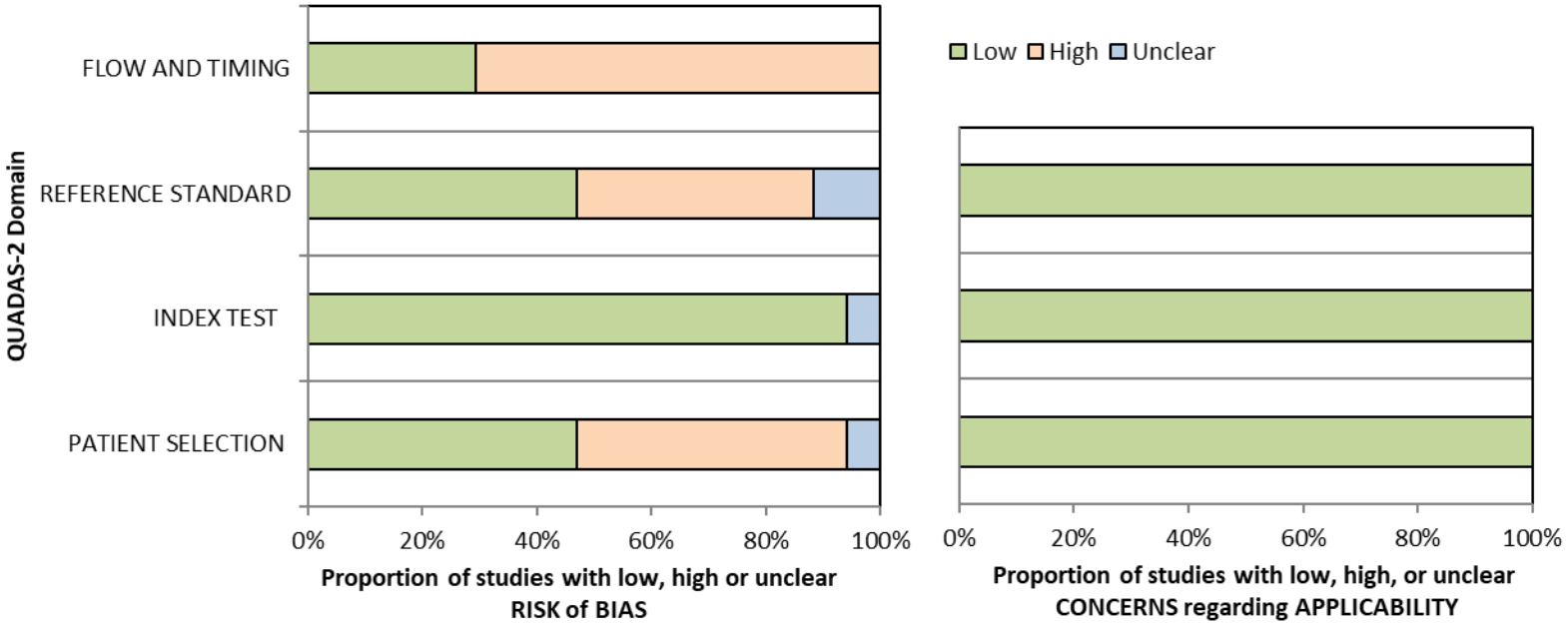


Table 10: QUADAS-2 tables for bone assessment methods to predict hip fracture

STUDY	RISK OF BIAS				APPLICABILITY		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Adami 2020	High	Low	High	Low	Low	Low	Low
Azagra 2016	Low	Low	Low	High	Low	Low	Low
Bolland 2011	Low	Low	Low	High	Low	Low	Low
Center 2004	Low	Low	Unclear	High	Low	Low	Low
Chan 2013	High	Low	Unclear	High	Low	Low	Low
Crandall 2020	High	Low	Low	High	Low	Low	Low
Cummings 1994	High	Low	Low	High	Low	Low	Low
Dargent-Molina 1999	High	Low	High	High	Low	Low	Low
Dargent-Molina 2003	Low	Low	High	Low	Low	Low	Low
Ensrud 2024	High	Unclear	Low	High	Low	Low	Low
Hans 2004	High	Low	Low	High	Low	Low	Low
Hillier 2007	Low	Low	Low	High	Low	Low	Low
Krieg 2006	Low	Low	Low	High	Low	Low	Low
Leslie 2007	Low	Low	High	Low	Low	Low	Low
Leslie 2014	High	Low	High	Low	Low	Low	Low
Schacter 2017B	Unclear	Low	High	Low	Low	Low	Low
Zarzour 2024	Low	Low	High	High	Low	Low	Low

Figure 11: QUADAS-2 domains for hip fracture



Appendix G Excluded studies

G.1 Accuracy of bone assessment methods for predicting fragility fracture

G.1.1 Clinical studies

Table 11: Studies excluded from the clinical review

Study	Code [Reason]
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
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	- Included in Evidence Review C
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*
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 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	- Systematic review used as source of primary studies *No relevant articles*

Study	Code [Reason]
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*
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

Study	Code [Reason]
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*
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 postmenopausal female population by binary statistical procedure CART. Revista de Osteoporosis y Metabolismo Mineral 12(4): 122-128	- Study design not relevant to this review protocol *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*
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	- Secondary publication of an included study that does not provide any additional relevant information

Study	Code [Reason]
density. Calcified tissue international 90(2): 128-36	*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).*
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*
Cheng, Xiaoguang, Wang, Ling, Wang, Qianqian et al. (2014). Validation of quantitative	- Study design not relevant to this review protocol

Study	Code [Reason]
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	*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; 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	- Included in Evidence Review C
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., 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*
Cummins, N M, Poku, E K, Towler, M R et al. (2011) clinical risk factors for osteoporosis in Ireland and the UK: a comparison of FRAX and QFractureScores. Calcified tissue international 89(2): 172-7	- Study design not relevant to this review protocol *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*

Study	Code [Reason]
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*
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 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	- Study does not contain an intervention relevant to this review protocol *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	- Comparator in study does not match that specified in this review protocol *Reference standard not described*

Study	Code [Reason]
densitometry : the official journal of the International Society for Clinical Densitometry 11(3): 397-403	
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*
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*
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	- Systematic review used as source of primary studies *Canadian review of FRAX and other models, UK studies not included in review*

Study	Code [Reason]
the accuracy of risk prediction tools. Systematic reviews 12(1): 51	
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 post-menopausal women. The British journal of radiology 80(959): 893-7	- Study design not relevant to this review protocol
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 osteoporosos/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	- Population not relevant to this review protocol *Population-based MrOS cohort*

Study	Code [Reason]
Cohort. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 38(9): 1258-1267	
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	- Included in Evidence Review C
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 chronic obstructive pulmonary disease. Chronic respiratory disease 16: 1479972318769763	- Study design not relevant to this review protocol *Cross-sectional study design*
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*
Hansen, KE, Blank, RD, 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	- Study design not relevant to this review protocol *Case-control study*

Study	Code [Reason]
National Osteoporosis Foundation of the USA 25(9): 2181-8	
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)*
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	- Included in Evidence Review C
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	- Included in Evidence Review C
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	- Included in Evidence Review C
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	- Included in Evidence Review C
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	- Included in Evidence Review C
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*

Study	Code [Reason]
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*
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*

Study	Code [Reason]
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 fractures: A systematic review and meta-analysis. Menopause: 1392	- Conference abstract
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*

Study	Code [Reason]
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 female patients with rheumatoid arthritis. Modern rheumatology 26(4): 540-5	- Data not reported in an extractable format or a format that can be analysed *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 :	- Study design not relevant to this review protocol *Cross-sectional diagnostic study*

Study	Code [Reason]
the official journal of the International Society for Clinical Densitometry 23(3): 373-380	
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*
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.; 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*
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	- Study design not relevant to this review protocol

Study	Code [Reason]
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	*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 femur in the prediction of osteoporotic spine fractures. European radiology 12(2): 401-8	- Study design not relevant to this review protocol *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*
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	- Included in Evidence Review C
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	- Included in Evidence Review C
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*

Study	Code [Reason]
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 early stage breast cancer patients treated with aromatase-inhibitors: An enhanced screening approach incorporating trabecular bone score. Journal of bone oncology 7: 32-37	- Study design not relevant to this review protocol *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	- Review article but not a systematic review

Study	Code [Reason]
FRAX model. Climacteric : the journal of the International Menopause Society 25(1): 22-28	
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 journal of the American Society for Bone and Mineral Research 35(3): 446-459	- Systematic review used as source of primary studies *No additional studies identified*
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*

Study	Code [Reason]
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*
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	- Study design not relevant to this review protocol

Study	Code [Reason]
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	*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*
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*
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	- Systematic review used as source of primary studies *No additional studies identified*

Study	Code [Reason]
review. Turkish journal of urology 45(4): 245-253	
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
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*

Study	Code [Reason]
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*
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*

Study	Code [Reason]
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 population in North-East Italy. Minerva ginecologica 66(5): 447-53	- Study design not relevant to this review protocol *Non-UK FRAX retrospective cohort study*
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. 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*
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	- Data not reported in an extractable format or a format that can be analysed *vertebral fractures reported*

Study	Code [Reason]
Efficacious in Predicting Thoracolumbar Fragility Fractures?. Neurospine 20(4): 1193-1204	
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*

Appendix H Research Recommendations

H.1 Recommendation for research

What is the accuracy of radiofrequency echo-graphic multi spectrometry (REMS) for predicting fragility fractures in adults, including those who have had a previous fragility fracture?

H.1.1 Why this is important

The measurement of BMD is an important part of the decision when assessing fracture risk and making decisions about whether treatment is clinically appropriate. It is also useful to monitor risk and treatment effectiveness. REMS is a more recent ultrasound-based and non-ionising radiation imaging technology used to predict fracture. The benefits of REMS are that is quick and safe. Further studies are needed to assess the ability of REMS to predict fragility fracture because it reduces exposure to ionising radiation and could increase accessibility to imaging services.

Importance to 'patients' or the population	More accurate tools for prediction of fragility fracture in adults would benefit patients through prevention, early intervention, and reduction of unnecessary treatment. It would lead to less ionising radiation exposure.
Relevance to NICE guidance	High. The research is essential to inform future updates as imaging methods using non-ionising radiation are to be preferred on patient safety and health grounds.
Relevance to the NHS	The NHS and commissioners of services would need to consider training staff costs, as well as the subsequent effects on DXA imaging services.
National priorities	Yes. The UK government has committed to addressing inequalities and a shift to prevention, and reducing time to 18 weeks wait for elective care https://www.england.nhs.uk/long-read/2025-26-priorities-and-operational-planning-guidance/
Current evidence base	The current evidence is limited to one small study, which has some methodological limitations.
Equality considerations	Intervention is non-ionising radiation imaging technology, which is safer for patients (in particular, pregnant people) and portable, which may expand accessibility in the community.

H.1.2 Modified PICOTS table

Population	Adults (18 years and older) at risk or suspected risk of fracture
Target condition	Major osteoporotic fracture; Hip fracture
Index test	Radiofrequency Echo-graphic Multi Spectrometry (REMS)
Reference standard	Expert radiological assessment of X-ray
Statistical measures	Accuracy of estimation of vertebral fracture: • Sensitivity/ specificity • Positive and negative likelihood ratio • Area under curve (AUC)/c-statistic
Study design	Diagnostic: cohort and cross-sectional studies
Timeframe	Completed prior to future updates of the osteoporosis guideline to inform future recommendations.