UK National Screening Committee

Screening for osteoporosis in postmenopausal women

External review against programme appraisal criteria for the UK National Screening Committee

Version: FINAL

Author: Solutions for Public Health

Date: January 2019

The UK National Screening Committee secretariat is hosted by Public Health England.

Template v17.0, July 2018

About the UK National Screening Committee (UK NSC)

The UK NSC advises ministers and the NHS in the 4 UK countries about all aspects of population screening and supports implementation of screening programmes.
Conditions are reviewed against evidence review criteria according to the UK NSC's evidence review process.

Read a complete list of UK NSC recommendations.

UK NSC, Floor 5, Wellington House, 133-155 Waterloo Road, London, SE1 8UG
www.gov.uk/uknsc
Twitter: @PHE_Screening Blog: phescreening.blog.gov.uk

For queries relating to this document, please contact:
phe.screeninghelpdesk@nhs.net

Published January 2019

Contents

Plain English summary

Osteoporosis is a condition that causes loss of bone strength. This tends to get worse as people get older. It is most common in women who have gone through the menopause (postmenopausal). If osteoporosis is not treated it can cause fractures in fragile bones from minor pressures. These are called fragility fractures. The most common fragility fractures occur in the spine, hip and wrist. They can reduce a person's quality of life and ability to live independently.

Osteoporosis is usually first found after a person has suffered a fragility fracture. In the UK there are about 536,000 new fragility fractures each year with high costs to the NHS. There is no treatment that can stop osteoporosis but some medicines can slow its progress. People with osteoporosis can also receive advice about how to reduce their risk of a fragility fracture. This is usually by lifestyle changes such as increasing exercise and avoiding falls.

This document looks at screening postmenopausal women for osteoporosis. It considers new evidence published between January 2011 and September 2018. A national screening programme would aim to prevent fragility fractures resulting from osteoporosis. It would also aim to help women maintain their independence and quality of life.

The UK NSC published its last review in 2013. This recommended against introducing a population screening programme for osteoporosis in postmenopausal women in the UK. The current review looked at some key questions:

1. how accurate are screening tests for osteoporosis?
2. how effective are treatments and changes in lifestyle in preventing fragility fractures caused by osteoporosis?
3. does screening reduce fractures caused by osteoporosis compared to usual care?
4. have studies shown that screening for osteoporosis is cost-effective in the UK?

The UK NSC still cannot recommend population screening for osteoporosis in postmenopausal women. There was not enough new

evidence to change the conclusions of the previous UK NSC review. These areas are still uncertain:

- the accuracy of screening tests in women who would be included in a population screening programme
- the effect of treatment and changes in lifestyle on some types of fracture
- the effect of treatment and changes in lifestyle in women identified as being at risk of fracture through screening
- how much added benefit would be gained by population screening over usual care
- the cost-effectiveness of a population screening programme.

Executive summary

Purpose of the review

This document reviews the evidence on population screening for osteoporosis in postmenopausal women.

Background

Osteoporosis is a skeletal disorder that tends to worsen with age and is most common in postmenopausal women. Without treatment, osteoporosis causes loss of bone mass which can increase the risk of fragility fractures from minor external pressure. These most commonly occur in the vertebra (spine), proximal femur (hip) and distal radius (wrist). Osteoporosis is usually detected after a person has suffered a fragility fracture. In the UK there are approximately 536,000 new fragility fractures each year with high associated costs to the NHS. The pain and loss of independence associated with osteoporotic fractures can reduce quality of life, with 50% of people suffering a hip fracture ceasing to live independently and 20% dying within 1 year of the fracture.

Risk assessment tools that calculate the 10-year probability of a fragility fracture can be used to assess people with suspected osteoporosis. A dual energy X-ray scan (DEXA) can be used to measure bone density. There is no cure for osteoporosis but treatment can slow progression and people can receive advice on minimising fracture risk by preventing falls and keeping active.

Focus of the review

The aim of a national screening programme targeting postmenopausal women would be to prevent osteoporotic fractures which would thereafter maintain independence and quality of life.

The population of interest for a national screening programme is postmenopausal women who have not had a diagnosis of osteoporosis or previously been identified at risk. The management of people already diagnosed with osteoporosis or who have suffered a clinically apparent osteoporotic fragility fracture prior to screening is outside the scope of this review. Guidance on the assessment, diagnosis and management of

osteoporosis is available from the National Institute of Health and Care Excellence (NICE) and the National Osteoporosis Guideline Group (NOGG).

This evidence summary includes studies published from January 2011 up to September 2018. It considers 4 key questions:

1. what is the accuracy of screening tests for osteoporosis?
2. what is the effectiveness of interventions in reducing the risk of osteoporotic fracture?
3. have RCTs demonstrated the clinical benefit of screening in reducing osteoporotic fractures in comparison to standard care?
4. have UK evaluations demonstrated that screening for osteoporosis is cost-effective?

Recommendation under review

The current UK NSC policy is that systematic population screening for osteoporosis in postmenopausal women is not recommended. The previous UK NSC external review of screening for osteoporosis was published in 2013. The 2013 review concluded that it was not appropriate to implement a national screening programme for osteoporosis. This was because, at that time, there were a number of uncertainties in the evidence base relating to screening tests; intervention in screen-detected populations and who to treat. There were also no RCTs assessing the clinical effectiveness of screening and treatment that were relevant to the UK. In addition the cost-effectiveness of screening had not been evaluated.

Findings and gaps in the evidence of this review

The current review found that there have been changes to the evidence base since the previous review, eg in the publication of 2 RCTs comparing screening to usual care. However the volume, quality and direction of new evidence published up to September 2018 is insufficient to change the conclusions of the previous UK NSC review. Remaining areas of uncertainty are:

- there are different potential approaches to screening for osteoporosis or risk of different types of fracture with varying test performance and a lack of studies in the specific population of interest
- concerns remain about the limited impact of pharmacological intervention on non-vertebral fractures and uncertainty about whether the evidence on the effectiveness of intervention is generalisable to a screened population

- there is an absence of studies demonstrating that screening has an impact on fracture-related morbidity or mortality
- there is a lack of evidence demonstrating an advantage to population screening over usual care from randomised controlled trials. However, the reduction in hip fracture observed as a secondary outcome in 1 RCT may warrant further research. Hip fractures can have long term health consequences, increase the risk of other (physical) problems, and so have negative impacts on quality of life
- the opportunity cost of a full population screening programme is uncertain.

Recommendations on screening

The current recommendation not to introduce a UK systemic population screening programme for osteoporosis in postmenopausal women should be retained.

Limitations

This rapid review process was conducted over a condensed period of time and did not include grey literature sources. Studies not available in the English language, abstracts and poster presentations, were not included. Studies that were not published in peer-reviewed journals were not reviewed.

Evidence uncertainties

There are a number of different potential approaches to screening identified within the studies included in the review and a lack of studies that meet all of the inclusion criteria for this review for some of the key questions. Further consideration and agreement concerning the focus of a population screening programme (eg osteoporosis or risk of any or specific fragility fractures), would help direct future research and reviews. There is also uncertainty about how much additional value would be gained by population screening over the current usual care.

Introduction and approach

Background

Osteoporosis is a systematic skeletal disorder that reduces bone mass as measured by bone mineral density (BMD) and has a tendency to worsen with age[1;2]. Without treatment, the continuing loss of bone mass can increase the risk of fragility fractures from minor external pressure[1]. Osteoporosis develops over several years and is generally painless until a fracture occurs[1]. The most common fragility fractures occur in the vertebra (spine), proximal femur (hip) and distal radius (wrist)[3]. Osteoporotic fractures can reduce quality of life due to chronic pain and loss of independence, with 50% of people suffering a hip fracture ceasing to live independently and 20% dying within 1 year of the fracture[1].

More than one-third of adult women and 1 in 5 men will sustain 1 or more osteoporotic fractures in their lifetime[2]. The female hormone oestrogen is essential for maintaining bone mineral density and when levels drop steeply after menopause; women are at increasing risk of developing osteoporosis. In men the male hormone testosterone is similarly important for maintenance of bone mineral density and although levels do drop gradually with age the risk of developing the condition is not as great in men as in women[4].

Osteoporosis is usually detected through the examination of a patient who has suffered a fracture from a minor fall or other sudden impact to the body1. In the UK there are approximately 536,000 new fragility fractures each year[5]. The previous review for the UK National Screening Committee (UK NSC) in 2013 found that osteoporosis-related fractures cost the NHS over £1.7 billion per year, with hip fractures accounting for most of the cost[1].

People with suspected osteoporosis are usually assessed using a risk assessment tool that calculates the 10-year probability of a fragility fracture occurring such as the Fracture Risk Assessment Tool (FRAX) or QFracture[1]. Clinical risk factors that are used in these risk assessment tools include: age, BMD, low body mass index, history of prior fracture at a site characteristic for osteoporosis, parental history of hip fracture, smoking, use of glucocorticoids, alcohol intake of ≥ 3 units per day, rheumatoid arthritis and diabetes[5].

A dual energy X-ray scan (DEXA) can be used to measure bone density[1]. The National Institute for Health and Care Excellence (NICE) recommend that people with a FRAX or Q- score close to the threshold for treatment should be considered for DEXA prior to initiation of treatment. NICE have also recommended research to explore the clinical utility of this strategy[1].

Osteoporosis is defined by a BMD that is 2.5 standard deviations or more below the young adult mean value for women (T-score less than or equal to -2.5 SD)[2]. Osteopenia or non-osteoporotic low bone density or mass is defined by a BMD between 1.0 and 2.5 standard deviations below the mean[2].

There is no cure for osteoporosis but treatment can slow progression and people can receive advice on minimising fracture risk by preventing falls and keeping active[1]. The NICE clinical guideline on assessing the risk of fragility fracture in osteoporosis was published in August 2012 and last updated in February 2017[3]. This recommends that assessment of fracture risk should be considered in all women aged 65 years and over, all men aged 75 years and over and, where risk factors are present, in women aged under 65 years and men aged under 75 years[3]. NICE have also published guidance on pharmacological interventions for osteoporosis. This includes the use of bisphosphonates (alendronate, etidronate, risedronate), a selective oestrogen receptor modulator (SERMS) (raloxifene), a calcium regulating drug (strontium ranelate) and a bone metabolism regulator (denosumab) for the prevention of osteoporotic fractures in postmenopausal women[6;7]. Guidance from the National Osteoporosis Guideline Group (NOGG) on the assessment, diagnosis and management of osteoporosis was updated in March 2017[5]. This recommends a case-finding strategy in the absence of population-based screening. In this case-finding strategy patients are identified because of a fragility fracture or by the presence of other clinical risk factors such as low body mass index, parental history of hip fracture, smoking, use of glucocorticoids, drinking 3 or more units of alcohol per day and rheumatoid arthritis[5].

The aim of a national screening programme targeting postmenopausal women would be to prevent osteoporotic fractures which would thereafter maintain independence and quality of life[1]. Approximately half of all hip fractures occur in postmenopausal women who have osteoporosis but no previous history of a clinically apparent osteoporotic fragility fracture[2]. The management of people already diagnosed with osteoporosis or who have suffered a clinically apparent osteoporotic fragility fracture prior to screening is outside the scope of this review.

Current policy context and previous reviews

The current UK NSC policy is that systematic population screening for osteoporosis in postmenopausal women is not recommended. The previous UK NSC external review of screening for osteoporosis was published in 2013 and was based on a literature search up to June 2011. The review concluded that it is not appropriate to implement a national screening programme for osteoporosis. The conclusions of the previous review are summarised below[1]:

- there had been no randomised controlled trials assessing the clinical effectiveness of screening and treatment which were relevant to current standards of care in the UK
- the cost-effectiveness of screening had not been evaluated.
- studies of DEXA alone had demonstrated poor sensitivity and a UK based RCT of screening using DEXA in combination with the FRAX risk assessment tool had not reported at the time of the 2013 review
- there were no studies exploring the frequency of screening
- there was concern about the limited impact of pharmacological treatment on non-vertebral fractures, that the evidence of these interventions was not directly generalisable to a screened population and that the duration of treatment appeared uncertain from the available evidence
- there was a lack of consensus between 2 leading sources of guidance in the UK regarding which women should be offered treatment
- a screening programme aiming to prevent osteoporotic fractures would not address the majority of fractures in postmenopausal women
- there were concerns that the DEXA capacity required to support a screening service was not available and that service expansion would be required
- although osteoporosis would be the target of a screening programme, the identification of women with osteopenia might create pressure to expand the screening programme. There had been no systematic review on the effectiveness of treatment and management options in women with reduced BMD meeting the criteria of osteopenia.

Objectives

The aim of the current review is to update the evidence in key areas identified in the previous review. The key questions addressed in the current review were developed by the UK NSC with input from Solutions for Public Health.

The key questions and the UK NSC criteria that they relate to are presented in Table 1 below.

Table 1. Key questions for the evidence summary, and relationship to UK NSC screening criteria

	Criterion	Key questions	Studies included
	THE TEST		
4	There should be a simple, safe, precise and validated screening test.	1. What is the accuracy of screening tests for osteoporosis?	1
5	The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed.		
	THE INTERVENTION		
9	There should be an effective intervention for patients identified through screening, with evidence that intervention at a pre-symptomatic phase leads to better outcomes for the screened individual compared with usual care. Evidence relating to wider benefits of screening, for example those relating to family members, should be taken into account where available. However, where there is no prospect of benefit for the individual screened then the screening programme shouldn't be further considered.	2. What is the effectiveness of interventions in reducing the risk of osteoporotic fracture?	1
10	There should be agreed evidence based policies covering which individuals should be offered interventions and the appropriate intervention to be offered.		
	THE SCREENING PROGRAMME		
11	There should be evidence from high quality randomised controlled trials that the screening programme is effective in reducing mortality or morbidity. Where screening is aimed solely at providing information to allow the person being screened to make an "informed choice" (eg. Down's syndrome, cystic fibrosis carrier screening), there must be evidence from high quality trials that the test accurately measures risk. The information that is provided about the test and its outcome must be of value and readily understood by the individual being screened.	3. Have RCTs demonstrated the clinical benefit of screening in reducing osteoporotic fractures in comparison to standard care?	2
12	There should be evidence that the complete screening programme (test, diagnostic procedures, treatment/ intervention) is clinically, socially and ethically acceptable to health professionals and the public.		
13	The benefit gained by individuals from the screening programme should outweigh any harms for example from overdiagnosis, overtreatment, false positives, false reassurance, uncertain findings and complications.		
14	The opportunity cost of the screening programme (including testing, diagnosis and treatment,	4. Have UK evaluations	1

	Criterion	Key questions	Studies included
	administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (ie. value for money). Assessment against this criteria should have regard to evidence from cost benefit and/or cost-effectiveness analyses and have regard to the effective use of available resource.	demonstrated that screening for osteoporosis is cost-effective?	

Methods

The current review was conducted by Solutions for Public Health (SPH), in keeping with the UK National Screening Committee evidence review process. Database searches were conducted on 13th February 2017 and 6th September 2018 to identify studies relevant to the questions detailed in Table 1. The search was conducted in 2 phases:

- a search in February 2017 included evidence published since January 2011. This searched for evidence relating to questions 1,2 and 3
- a search in September 2018 included evidence published since February 2017 for questions 1,2, and 3, and evidence published since January 2011 for question 4.

The second search was conducted as two randomised controlled trials of screening had been published after the first search date.

Eligibility for inclusion in the review

The following review process was followed:

1. each abstract was reviewed against the inclusion/exclusion criteria by 1 reviewer. Where the applicability of the inclusion criteria was unclear, the article was included at this stage in order to ensure that all potentially relevant studies were captured.
2. full-text articles required for the full-text review stage were acquired.
3. each full-text article was reviewed against the inclusion/exclusion criteria by 1 reviewer, who determined whether the article was relevant to 1 or more of the review questions.
4. any queries at the abstract or full-text stage were resolved through discussion with a second reviewer.
5. the review was quality assured by a second senior reviewer, not involved with the writing of the review in accordance with SPH's quality assurance process.

Eligibility criteria for each question are presented in Table 2 below.

The February 2017 search generated a total of 3,768 unique references which were sifted by title and abstract by an information scientist for potential relevance to the review. An SPH reviewer assessed 715 titles and abstracts for further appraisal and possible inclusion in the final review.

The September 2018 search generated a total of 938 unique references which were sifted by title and abstract by an information scientist for potential relevance to the review. An SPH reviewer assessed 65 titles and abstracts for further appraisal and possible inclusion in the final review.

Overall, 57 studies from the first search and 20 from the second search were identified as possibly relevant during title and abstract sifting and further assessed at full text (see Appendix 2 for study flow).

Table 2. Inclusion and exclusion criteria for the key questions

Key question	Inclusion criteria							Exclusion criteria
	Population	**Target condition**	**Intervention**	**Reference Standard**	**Comparator**	**Outcome**	**Study type**	
1. What is the accuracy of screening tests for osteoporosis?	Postmenopausal women (screened populations or comparable populations eg osteoporotic women with no history of fracture)	Osteoporosis	Any combination of fracture risk assessment + DEXA	Fragility fractures eg vertebral, hip confirmed by clinical assessment	Dependent on the intervention (eg risk assessment alone, standard practice as defined by protocol) No comparator	Sensitivity Specificity Area under the curve Positive predictive value Negative predictive value	Prospective studies of consecutively enrolled women should be prioritised	None stated
2. What is the effectiveness of interventions in reducing the risk of osteoporotic fracture?	Postmenopausal women (screened populations or comparable populations eg osteoporotic women with no history of fracture)	Osteoporosis	Exercise (load bearing, cardiovascular) Pharmacological interventions (eg bisphosphonates, denosumab, vitamin D supplementation, calcium supplementation etc)	N/a	In studies of screening, women identified through standard care	Osteoporotic fracture (vertebral, non vertebral, hip) All cause fracture Mortality Harms of treatment (thromboembolic event, osteonecrosis of the jaw, gastrointestinal problems)	RCTs and other studies in screen-detected populations should be prioritised	None stated

3. Have RCTs demonstrated the clinical benefit of screening in reducing osteoporotic fractures in comparison to standard care?	Postmenopausal women	Osteoporosis	Pharmacological and/or other interventions in women detected by screening based on risk assessment + DEXA	N/a	Pharmacological and/or other interventions in women detected by standard care	Osteoporotic fractures All clinical fractures Mortality Screening uptake rate Treatment compliance rate Physical/ psychological health state measures	RCTs	None stated
4. Have UK evaluations demonstrated that screening for osteoporosis is cost-effective?	Postmenopausal women	Osteoporosis	Screening – any combination of fracture risk assessment + DEXA	N/a	No screening Standard care	Cost-effectiveness measured by QALYs	Studies conducted in the UK which factor in recent RCT evidence: Cost-effectiveness analyses Technology assessments Systematic reviews and meta-analyses Modelling studies	None stated

Appraisal for quality/risk of bias tool

The following tools were used to assess the quality and risk of bias of each study included in the review:

- systematic reviews: Critical Appraisal Skills Programme (CASP) Systematic Reviews Checklist
- RCTs: Cochrane Collaboration's "Risk of Bias" Tool
- cost-effectiveness studies: CASP checklist for economic evaluations.

Results of the quality assessments are presented in the summary and appraisal of individual studies in Appendix 3.

Databases/sources searched

Systematic searches of 3 databases (Medline, Embase and Cochrane) were conducted on 13th February 2017 and 6th September 2018 to identify studies relevant to the questions detailed in Table 1. The search strategy is presented in Appendix 1.

Question level synthesis

Criteria 4 and 5

4. There should be a simple, safe, precise and validated screening test.

5. The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed.

Question 1 – What is the accuracy of screening tests for osteoporosis?

The previous 2013 UK NSC review[2] described 2 risk assessment tools to assess 10-year risk of fracture available for use in the UK (QFracture and FRAX). QFracture is a 17-item instrument that does not include a DEXA scan. The FRAX tool uses 12 items of clinical data to predict the 10-year probability of hip fracture or major osteoporotic fracture and can be used with or without a DEXA scan[2;8]. The 2013 review reported the area under the curve (AUC)* of QFracture as 0.86 to 0.89 for any fracture and the AUC of the FRAX tool as 0.54 to 0.78 for any osteoporotic fracture and 0.65 to 0.81 for hip fracture[2†].

Guidance from NICE recommends the use of FRAX or QFracture in the UK and notes that tools validated in other populations may not apply to the UK as their accuracy depends on the epidemiological data used in their design[3]. NICE also state that the measurement of BMD using DEXA should be considered in people whose fracture risk is in the region of an intervention threshold for a proposed treatment with FRAX. Risk should then be recalculated with the addition of the DEXA BMD value[3]. The latest guidance from the NOGG recommends that risk of fracture should be expressed as probability over a 10-year interval with FRAX as the preferred fracture risk assessment tool[5].

Because of the poor performance of DEXA alone reported in previous UK NSC reviews, the screening strategy of interest in the current review is any combination of fracture risk assessment plus DEXA[1].

* An area of 1 represents a perfect test, an area of 0.5 represents a test equally likely to produce false positive or true positive results

† Test performance was not presented in terms of sensitivity or specificity in the 2013 review or the systematic review that was the source of the information

Eligibility for inclusion in the review

The inclusion criteria for this question are summarised briefly below:

- population – postmenopausal women (screened populations or comparable populations eg osteoporotic women with no history of fracture)
- index test - any combination of fracture risk assessment + DEXA
- reference standard – fragility fractures eg vertebral, hip
- comparator - dependent on the intervention (eg risk assessment alone, standard practice as defined by protocol) or no comparator
- outcomes – sensitivity, specificity, area under the curve, positive predictive value, negative predictive value
- study design - prospective studies of consecutively enrolled women should be prioritised
- date and language – studies published in English after 1st January 2011

Description of the evidence

Database searches yielded 780 results, of which 151 were judged to be relevant to this question and 20 abstracts met the criteria for full text review. However few of these assessed the performance of a combination of fracture risk assessment plus DEXA and no studies met all of the inclusion criteria‡. After review of the full texts, 1 systematic review was included.

Reasons for excluding studies after review of the full text were:

- the population did not comprise postmenopausal women
- the population included people with prior fracture
- the population was assessed for osteoporosis opportunistically during other medical assessments/ procedures
- individual studies that only assessed the performance of a fracture risk assessment tool without BMD§
- studies focusing on tests not recommended for use in the UK eg studies considering risk assessment tools developed for non-UK populations
- studies focusing on the use of tests in non-UK populations eg the development of population specific thresholds
- studies using medical records to retrospectively calculate risk in referred populations
- studies comparing diagnostic tests (not screening tests assessing fracture risk)

‡ Two RCTs of screening have been completed. However the study publications do not provide information on test performance

§ Studies assessing performance without BMD, where BMD is optional for that tool, were included within the included systematic review

- studies focusing on prevalence rather than test performance
- studies focusing on aspects of screening outside of the key questions for this review
- case control studies (ie not a consecutively enrolled population)
- descriptive reviews/ commentary
- older version of a NOGG guideline that has since been updated.

Discussion of findings

A study-level summary of data extracted from each included publication is presented in the appraisal of individual studies in Appendix 3. In Appendix 3 publications are stratified by question.

No studies were identified that explored the test performance of screening using a combination of fracture risk assessment and DEXA in a population of consecutively enrolled postmenopausal women. The systematic review by Viswanathan et al (2018)[9] considered evidence published between November 2009 and March 2018 and provides an overview of the performance of a range of fracture risk assessment tools in combination with BMD measurement (eg DEXA). However, the results should be treated with some caution as the inclusion criteria for the systematic review were broader than those specified for this review and there was insufficient detail to determine whether the study populations were consecutively enrolled or the comparator/ reference standard used. The systematic review also included women aged 40 years or more rather than postmenopausal women specifically. However, the lower age range or mean age of women in the studies included in the systematic review tended to be over 50[8]. To meet the inclusion criteria for the systematic review, women had to be community-dwelling with no known low-trauma fractures or metabolic bone disease.

Table 3 summarises the test performance of 6 risk assessment tools combined with BMD measurement. For some tools the use of BMD is optional. Where reported, the performance of the tool without BMD is also provided for information.

The nature of the risk assessed varied, with several tools reporting performance for predicting more than 1 type of fracture. Across the 6 tools, risk was assessed for major osteoporotic fracture, any osteoporotic fracture, hip fracture, non-vertebral fracture and long bone and vertebral fracture. The timeframe also varied from 1 year to 10 years. The areas under the curve

reported for the risk assessment tools with BMD measurement ranged from 0.62 to 0.85. The studies included in the systematic review varied considerably in size. For the two tools with pooled results this ranged from approximately 6,000 to over 190,000 patients and between 3 and 17 studies. The performance of the other tools was assessed in a single study with sample sizes ranging from 400 to over 94,000.

Results for the FRAX tool were pooled from 12 studies and this is the only tool included in the systematic review that is recommended for use in the UK by NICE. FRAX can be used to predict either the 10-year risk of major osteoporotic fracture or hip fracture, achieving pooled AUC scores of 0.70 (95%CI 0.68 to 0.71) for major osteoporotic fracture (12 studies, n=62,054) and 0.79 (95%CI 0.76 to 0.81) for hip fracture (10 studies, n=161,984) when combined with BMD measurement. The only other tool using data from multiple studies was the Garvan nomogram/Fracture Risk Calculator. Pooled AUC scores were 0.68 (95%CI 0.64 to 0.71) for 10-year risk of major osteoporotic fracture (3 studies, n=6,174) and 0.72 (95%CI 0.66 to 0.79) for 10-year risk of hip fracture (4 studies, n=7,449).

The review was assessed using the CASP checklist for systematic reviews. There were no concerns about the conduct of the systematic review. The review authors assessed the quality of the included studies which were all judged to be of good or fair quality. Issues with the evidence base identified by the review authors included inconsistent reporting of the version of risk assessment tools used and studies with follow-up periods that were shorter than the timeframe covered by the risk assessment tool. A pooled area under the curve was calculated for tools with results from more than 1 study. Where reported the heterogeneity scores (I^2)** were very high indicating considerable variation between the studies included in the pooled analysis and reducing confidence in the results. The review authors noted that there was considerable heterogeneity in the patient populations and length of study follow-up in the included studies. Some studies included in the review did not report 95% confidence intervals and there was variation in the width of the confidence intervals across those that did.

** I^2 – is a measure of heterogeneity relating to the variation in study outcome between studies

Table 3. Summary of test performance for risk assessment tools from Viswanathan et al (2018)[9]

Assessment tool	Risk assessed		Area under the curve (AUC)††	Number of studies (n)
FRAX	10-year risk of major osteoporotic fracture	With BMD	0.70 (95%CI 0.68 to 0.71), I^2=92.1%	12 (n=62,054)
		Without BMD	0.66 (95%CI 0.63 to 0.69), I^2=99.2%	17 (n=158,897)
	10-year risk of hip fracture	With BMD	0.79 (95%CI 0.76 to 0.81), I^2=99.1%	10 (n=161,984)
		Without BMD	0.76 (95%CI 0.72 to 0.81), I^2=99.8%	12 (n=190,795)
Garvan nomogram/ Fracture Risk Calculator	10-year risk of major osteoporotic fracture	With BMD	0.68 (95%CI 0.64 to 0.71), I^2=84.8%	3 (n=6,174)
		Without BMD	0.66 (95%CI 0.61 to 0.72)	1 (n=600)
	10-year risk of any osteoporotic fracture	With BMD	0.69 (95%CI not reported)	1 (n=506)
		Without BMD	0.65 (95%CI not reported)	1 (n=506)
	10-year risk of hip fracture	With BMD	0.72 (95%CI 0.66 to 0.79), I^2=97.3%	4 (n=7,449)
		Without BMD	0.68 (95%CI not reported)	1 (n=1,369)
	10-year risk of non-vertebral fracture	With BMD	0.62 (95%CI not reported)	1 (n=1,646)
		Without BMD	0.58 (95%CI not reported)	1 (n=1,637)
Women's Health Initiative	5-year risk of hip fracture	With BMD	0.80 (95%CI 0.75 to 0.85)	1 (n=10,750)
		Without BMD	In 2 studies‡‡: 0.80 (95%CI 0.77 to 0.82) 0.82 (95% CI not reported)	1 (n=10,750) 1 (n=13,353)

†† Area under the curve was the test performance metric reported by the systematic review. Sensitivity and specificity were not provided
‡‡ Pooled AUC not reported

FRISC	1,3,5 or 10[§§]-year risk of major osteoporotic fracture	With BMD	0.73 (95%CI not reported)	1 (n=400)
	1,3,5 or 10-year risk of long bone and vertebral fracture	With BMD	0.69 (95%CI 0.64 to 0.73)	1 (n=765)
FRISK	5 or 10-year risk of major osteoporotic fracture[§§]	With BMD	0.66 (95%CI 0.60 to 0.71)	1 (n=600)
		Without BMD	0.62 (95%CI 0.56 to 0.67)	1 (n=600)
FRC	10-year risk of hip fracture	With BMD	0.85 (95%CI 0.84 to 0.86)	1 (n=94,489)
		Without BMD	0.83 (95%CI 0.82 to 0.84)	1 (n=94,489)

BMD – bone mineral density; FRAX – Fracture Risk Assessment Tool; FRC – Fracture risk calculator; FRISC – Fracture and Immobilization Score; FRISK - Fracture Risk Score

[§§]The tool can be used to assess risk over different timeframes. Timeframe of risk assessment for the reported study not specified

Summary of Findings Relevant to Criteria 4 and 5: Criteria not met***

The previous 2013 UK NSC review reported an AUC of 0.86 to 0.89 for any fracture for QFracture and 0.54 to 0.78 for FRAX for any osteoporotic fracture and 0.65 to 0.81 for hip fracture.

There was a lack of studies that met all of the inclusion criteria for this review. A 2018 systematic review conducted for the US Preventative Services Task Force provides an overview of the performance of 6 risk assessment tools with measurement of BMD, albeit with broader inclusion criteria than the current review.

The AUCs reported for risk assessment tools combined with BMD measurement ranged from 0.62 to 0.85. Where tools assessed risk of different types of fracture the AUC tended to be slightly higher for assessment of risk of hip fracture. An AUC of 1 represents a perfect test whereas an AUC of 0.5 represents a test that is equally likely to produce false positive or true positive results. There was a very high level of heterogeneity in the studies included in the pooled analysis reducing confidence in the results.

There are different potential approaches to screening for osteoporosis or risk of different types of fracture and a lack of studies that meet all of the inclusion criteria for this review. For this reason these criteria are currently not met.

Further consideration and agreement concerning the test approach to be used in a population screening programme (eg osteoporosis or risk of any or specific fragility fractures), would help direct future research and reviews to determine the performance of a specific approach in a consecutively enrolled population of postmenopausal women.

*** **Met** -for example, this should be applied in circumstances in which there is a sufficient volume of evidence of sufficient quality to judge an outcome or effect which is unlikely to be changed by further research or systematic review.
Not Met - for example, this should be applied in circumstances where there is insufficient evidence to clearly judge an outcome or effect or where there is sufficient evidence of poor performance.
Uncertain -for example, this should be applied in circumstances in which the constraints of an evidence summary prevent a reliable answer to the question. An example of this may be when the need for a systematic review and meta-analysis is identified by the rapid review.

Criteria 9 and 10

9. There should be an effective intervention for patients identified through screening, with evidence that intervention at a pre-symptomatic phase leads to better outcomes for the screened individual compared with usual care. Evidence relating to wider benefits of screening, for example those relating to family members, should be taken into account where available. However, where there is no prospect of benefit for the individual screened then the screening programme shouldn't be further considered.

10. There should be agreed evidence based policies covering which individuals should be offered interventions and the appropriate intervention to be offered.

Question 2 – What is the effectiveness of interventions in reducing the risk of osteoporotic fracture?

Sub question: Have any studies explored the effectiveness of interventions in reducing mortality?

The previous 2013 UK NSC review[2] did not identify any meta-analyses on the effect of treatments for osteoporosis in women detected through population-based screening. The 2013 review stated that ideally, such meta-analyses should only include women who have both osteoporosis and either no fractures or only subclinical vertebral fractures[2]. In the absence of 'ideal' data, the 2013 review referenced reviews by NICE[6;7] and Nelson et al (2010)[10] to summarise the evidence for the effectiveness of the major pharmacological interventions for the prevention of fracture. The Nelson review included trials that met 1 of the following criteria[2]:

- excluded individuals with previous vertebral or other presumably osteoporotic fractures
- permitted individuals with previous osteoporotic fractures, but the overall proportion of participants with fracture was <20%, or the trial reported results separately for participants with and without previous fractures
- did not report the proportion of participants with previous osteoporotic fractures, but the inclusion criteria did not select individuals on the basis of presence of a previous fracture, and mean BMD T-scores were ≥3.0.

Broadly, the previous 2013 UK NSC review found that across pharmacological interventions the reduction in vertebral fracture rate with treatment ranged from 40% to 70% with the reduction in non-vertebral fracture rate in the majority of studies ranging from 5% to 20%[2]. The 2013 review also identified a number of harms associated with treatment including thromboembolic events, osteonecrosis of the jaw and gastrointestinal problems[2]. The 2013 review did not identify any trial assessing the effect of fracture-related morbidity and mortality[2].

Eligibility for inclusion in the review

The inclusion criteria for this question are summarised briefly below:

- population – postmenopausal women (screened populations or comparable populations eg osteoporotic women with no history of fracture)
- intervention – exercise (load bearing, cardiovascular); pharmacological interventions (eg bisphosphonates, denosumab, vitamin D supplementation, calcium supplementation etc)
- comparator - in studies of screening, women identified through standard care
- outcomes – osteoporotic fracture (vertebral, non-vertebral, hip); all cause fracture; mortality; harms of treatment (thromboembolic event, osteonecrosis of the jaw, gastrointestinal problems)
- study design - RCTs and other studies in screen-detected populations should be prioritised
- date and language – studies published in English after 1st January 2011.

Description of the evidence

Database searches yielded 780 results, of which 579 were judged to be relevant to this question and 42 abstracts met the criteria for full text review.

A large number of studies relating to the treatment of osteoporosis were returned by the literature search. There is existing UK guidance on the treatment and management of osteoporosis from NICE[3] and NOGG[5]. The latest 2017 guideline from NOGG was accredited by NICE[5]. This update review is specifically concerned with evidence about the effectiveness of treatment in screen-detected or comparable populations of postmenopausal women and does not consider the wider evidence base for the effectiveness of different treatments or management strategies for osteoporosis.

After review of the full texts, 1 systematic review was included.

Reasons for excluding studies at this stage included:

- the population did not comprise postmenopausal women with osteoporosis
- the population included people with prior fractures
- studies did not include fracture or harm of treatment or mortality as an outcome (outcomes of interest for this review)
- studies focusing on monitoring treatment/ adherence to medication
- studies focusing on treatment of fracture (rather than treatment of osteoporosis)
- studies comparing treatments with each other/ equivalence (in non-screened populations)
- studies on treatment dosage
- studies on duration of treatment, length of effect or treatment discontinuation
- studies on effect of other medications on efficiency of bisphosphonates
- trial protocols (not-screening trials)
- non-randomised studies not using screen-detected populations
- studies focusing on clinical decision making
- studies on determining the mechanism by which treatments are effective
- studies on cost-effectiveness
- discussion/ commentary
- guidelines or consensus statements on management of osteoporosis from non-UK countries.

In addition, 2 RCTs of screening compared to usual care were identified (Rubin et al 2018[11]; Shepstone et al 2018[12]). These included the detection of postmenopausal women at risk of fracture and the recommendation that they be treated according to national guidelines. As no details were provided of the advice or treatment, if any, received by the women in either group (other than number of prescriptions for anti-osteoporotic medication) these studies do not provide any evidence about the effectiveness of intervention and are not included for this question. Further details of these studies can be found in the response to question 3.

Discussion of findings

A study-level summary of data extracted from each included publication is presented in the summary and appraisal of individual studies in Appendix 3. In Appendix 3 publications are stratified by question.

No studies exploring the effectiveness of interventions in screened populations were identified and no recent studies were identified in comparable populations of osteoporotic women with no history of fracture.

A US Preventative Services Task Force (USPSTF) systematic review on screening to prevent osteoporotic fracture (Viswanathan et al 2018)[9] has recently been published. This considered evidence published between November 2009 and March 2018 and is an update of the Nelson et al (2010) review used as a key source in the 2013 UK NSC review. The key question considered by Viswanathan et al (2018)[9] was 'what is the effectiveness of pharmacotherapy for the reduction of fractures and related morbidity and mortality?' The inclusion criteria for Viswanathan et al were studies where the majority of participants had an increased fracture risk with no reference to prior fracture. The Viswanathan et al systematic review included a broader study population (adults over 40 years old) however results for men and women were reported separately and examination of the titles of the intervention studies included in the review suggests that the majority of these focused on the effectiveness of intervention in postmenopausal women.

The results reported by Viswanathan et al are summarised in Table 4. For vertebral fractures, all pharmacological interventions reported were associated with a statistically significant reduction in fractures, with relative reduction in vertebral fracture rate ranging from 36% to 68%. For non-vertebral fractures, 2 interventions did not show a statistically significant improvement compared to placebo. However a statistically significant reduction in fractures was seen with bisphosphonates and denosumab, with reduction in non-vertebral fracture rate ranging from 16% to 20%. For hip fracture a statistically significant reduction was reported with denosumab (40%) in 1 study, but other interventions did not demonstrate a significant improvement compared to placebo.

Where reported, studies generally did not show any significant differences between intervention and placebo in serious adverse events, discontinuation of treatment, or specific adverse events such as upper gastrointestinal events, coronary heart disease or stroke. Oestrogen, with or without progesterone was associated with higher rates of various adverse events but limited details were reported.

The studies did not report fracture-related morbidity or mortality as outcomes.

The review was assessed using the CASP checklist for systematic reviews. There were no concerns about the conduct of the systematic review. The review authors assessed the quality of the included RCTs which were all judged to be of good or fair quality. Pooled analysis of multiple studies was performed for the outcomes of the use of bisphosphonates in vertebral (5 RCTs, n=5,433), non-vertebral (8 RCTS, n=16,438) and hip fractures (3 RCTs, n=8988). In all cases the heterogeneity between studies was reported as 0% suggesting consistency between the different studies. For all other interventions results were from single trials, and the study authors noted that outcomes were often dominated by a single large trial for each drug. The confidence intervals around many of the outcomes were wide reducing confidence in the results.

Table 4. Summary of the effectiveness of pharmacological interventions compared to placebo from Viswanathan et al (2018)[9]

Fracture site	Intervention	Outcome	Relative Risk	Number of studies (n)
Vertebral fracture	Bisphosphonates	Intervention: 2.1% Placebo: 3.8%	0.57 (95%CI 0.41 to 0.78), I^2=0%	5 RCTS (n=5,433)
	Denosumab	Intervention: 2.3% Placebo: 7.2%	0.32 (95%CI 0.26 to 0.41)	1 RCT (n=7,868)
Radiographic vertebral fracture	Raloxifene	Intervention: 7.5% Placebo: 12.5%	0.64 (95%CI 0.53 to 0.76)	1 RCT (n=7,705)
	Parathyroid hormone	Intervention: 0.7% Placebo: 2.1%	0.32 (95%CI 0.14 to 0.75)	1 RCT (n=2,061)
Non-vertebral fracture	Bisphosphonates	Intervention: 8.9% Placebo: 10.6%	0.84 (95%CI 0.76 to 0.92), I^2=0%	8 RCTS (n=16,438)
	Raloxifene	No significant difference between intervention and placebo (details not reported)	N/a	Not reported
	Denosumab	Intervention: 6.1% Placebo: 7.5%	0.80 (95%CI 0.67 to 0.95)	1 RCT (n=7,868)
	Parathyroid hormone	Intervention: 5.6% Placebo: 5.8%	0.97 (95%CI 0.71 to 1.33)	1 RCT (n=2,532)
Hip fracture	Bisphosphonates	Intervention: 0.7% Placebo: 0.96%	0.70††† (95%CI 0.44 to 1.11), I^2=0%	3 RCTS (n=8,988)
	Raloxifene	No significant difference between intervention and placebo (details not reported)	N/a	Not reported
	Denosumab	Intervention: 0.7% Placebo: 1.1%	0.60 (95%CI 0.37 to 0.97)	1 RCT (n=7,868)
Osteoporotic fracture	Oestrogen	Lower risk compared to placebo (figures not reported)	Hazard ratio: 0.72 (95%CI 0.64 to 0.80)	1 SR (n not reported)
Fracture	Oestrogen plus progestin	Lower risk compared to placebo (figures not reported)	0.80 (95%CI 0.68 to 0.94)	1 SR (n not reported)

RCT – randomised controlled trial; SR – systematic review

††† The systematic review authors reported that only 1 of the 3 RCTs was sufficiently powered to detect differences in hip fracture

Summary of Findings Relevant to Criteria 9 and 10: Criteria not met‡‡‡

As with the previous 2013 UK NSC review, the current review found a lack of evidence on the effects of treatments for osteoporosis in women detected through population-based screening.

The relative risk reductions identified by the 2018 systematic review were similar to the results reported in the previous 2013 UK NSC review, although with a narrower range of results for non-vertebral fractures. For vertebral fracture this was 36% to 68% compared to 40% to 70% and for non-vertebral fracture this was 16% to 20% compared to 5 to 20%. The studies did not report fracture-related morbidity or mortality as outcomes.

There is UK guidance on the treatment and management of osteoporosis. However, this is not specific to individuals identified through screening. RCTs have compared the effectiveness of screening to usual care (see question 3) but these did not provide full details of what, if any, advice or treatment women received.

There is insufficient new evidence to change the conclusions of the previous review about the effectiveness of interventions in reducing the risk of osteoporotic fracture in a screen-detected population. Therefore these criteria are not met.

‡‡‡ **Met** -for example, this should be applied in circumstances in which there is a sufficient volume of evidence of sufficient quality to judge an outcome or effect which is unlikely to be changed by further research or systematic review.
Not Met - for example, this should be applied in circumstances where there is insufficient evidence to clearly judge an outcome or effect or where there is sufficient evidence of poor performance.
Uncertain -for example, this should be applied in circumstances in which the constraints of an evidence summary prevent a reliable answer to the question. An example of this may be when the need for a systematic review and meta-analysis is identified by the rapid review.

Criteria 11, 12 and 13

11. There should be evidence from high quality randomised controlled trials that the screening programme is effective in reducing mortality or morbidity. Where screening is aimed solely at providing information to allow the person being screened to make an "informed choice" (eg. Down's syndrome, cystic fibrosis carrier screening), there must be evidence from high quality trials that the test accurately measures risk. The information that is provided about the test and its outcome must be of value and readily understood by the individual being screened.

12. There should be evidence that the complete screening programme (test, diagnostic procedures, treatment/ intervention) is clinically, socially and ethically acceptable to health professionals and the public.

13. The benefit gained by individuals from the screening programme should outweigh any harms for example from over diagnosis, overtreatment, false positives, false reassurance, uncertain findings and complications.

Question 3 – Have RCTs demonstrated the clinical benefit of screening in reducing osteoporotic fractures in comparison to standard care?

The previous 2013 UK NSC review did not identify any RCTs assessing the effectiveness of an osteoporosis screening programme on reducing mortality[2]. The 2013 review did identify RCTs on the effectiveness of osteoporosis screening on reducing fracture incidence but concluded that the results of these RCTs were not applicable as recommended treatments had changed since they were conducted[2].

Eligibility for inclusion in the review

The inclusion criteria for this question are summarised briefly below:

- population – postmenopausal women
- intervention – pharmacological and/or other interventions in women detected by screening based on risk assessment + DEXA
- comparator – pharmacological and/or other interventions in women detected by standard care
- outcomes – osteoporotic fractures; all clinical fractures; mortality; screening uptake rate; treatment compliance rate; physical/ psychological health state measures

- study design – RCTs
- date and language – studies published in English after 1st January 2011.

Description of the evidence

Database searches yielded 780 results, of which 43 were judged to be relevant to this question and 14 abstracts met the criteria for full text review. After review of the full texts, 2 studies were included.

Reasons for excluding studies after review of the full text included:

- studies about the implementation/ uptake of screening in the United States
- qualitative studies focusing on attitudes to screening
- reviews where the included studies did not meet the PICO (eg due to study population or design)
- a systematic review that included 1 of the RCTs separately identified for this review with limited details about the study
- a post-hoc analysis examining factors which might influence screening outcomes, but using FRAX scores without BMD to predict risk
- a study focusing for reasons for non-participation in a Danish screening trial.

Discussion of findings

A study-level summary of data extracted from each included publication is presented in the summary and appraisal of individual studies in Appendix 3. In Appendix 3 publications are stratified by question.

Two RCTs were identified comparing screening to usual care; the SCOOP trial (Shepstone et al 2018[12]) set in the UK and the ROSE trial (Rubin et al 2018[11]) set in Denmark. The key features and results of these studies are summarised in Table 5. Further details of these studies are provided in Appendix 3.

Table 5. Summary of the SCOOP and ROSE trials

	SCOOP[12]	ROSE[11]
Population	12,495[§§§] women aged 70 to 85 years living in 7 UK regions who had consented to participate in the RCT. 5 year follow-up	34,229 women aged 65 to 80 living in Southern Denmark who were invited to take part in the screening RCT[****]. 5 year follow-up
Screening group	• n=6,233 • completed baseline questionnaire to capture FRAX risk factors • women at high risk of hip fracture (10-year risk above an age-dependent threshold) offered a DEXA scan (n=3,064; 49%) • FRAX plus DEXA used to re-calculate 10-year risk of hip fracture • Mean femoral neck T score was -2.6 for high risk women • women and GP advised of the results • women at high risk advised to make an appointment with their GP (n=898; 14%)	• n=9,279 • completed baseline questionnaire to capture FRAX risk factors • women at moderate to high 10-year risk of major osteoporotic fracture (10 year probability ≥15%) offered a DEXA scan (n=6,226; 67%) • osteoporosis diagnosed if any sites measured on DEXA had a T score of ≤ -2.5 • women and GP advised of the results. GPs received information on national guidelines • recommended for treatment n=1,236 (13%)
Usual care group	• n=6,250 • completed baseline questionnaire to capture FRAX risk factors • FRAX risk not calculated until the end of the trial • GP notified of participation in the study only	• N=9,326 • completed baseline questionnaire to capture FRAX risk factors • women were not informed about the result of their FRAX calculation
Intervention (where reported)	Received ≥1 prescription for anti-osteoporotic medication during the study • screening group: 1,486 (24%) • usual care: 982 (16%) (no significance test reported)	Received osteoporotic medication during the study • screening group: 23% (n not specified) • usual care: 18% (n not specified) (p<0.001)
Analysis	Intention-to-treat	Intention-to-treat Pre-specified per-protocol

[§§§] 12 patients (6 from each group) were excluded after randomisation

[****] 18,605 women (54.3%) returned sufficient data to calculate FRAX and were not already receiving osteoporosis treatment

	SCOOP[12]	ROSE[11]
Osteoporosis-related fracture†††† (primary outcome)	No significant difference between screening (12.9%) and usual care (13.6%) (HR 0.94 95%CI 0.85 to 1.03, p=0.178)	N/a
Major osteoporotic fracture‡‡‡‡ (primary outcome)	N/a	Intention-to- treat analysis: No significant difference between screening (9.9%) and usual care (10.0%) (SHR§§§§ 0.986 95%CI 0.92 to 1.06, p=0.68) Pre-specified per-protocol analysis: No significant difference between screening (7.8%) and usual care (8.4%) (SHR 0.914 95%CI 0.83 to 1.01, p=0.08)
Any fracture (secondary outcome)	No significant difference between screening (15.3%) and usual care (16.0%) (HR 0.94 95%CI 0.86 to 1.03, p=0.183)	Intention-to- treat analysis: No significant difference between screening (13.1%) and usual care (13.0%) (SHR 1.004 95%CI 0.94 to 1.06, p=0.91) Pre-specified per-protocol analysis: No significant difference between screening (10.7%) and usual care (11.0%) (SHR 0.968 95%CI 0.89 to 1.06, p=0.47)
Hip fracture (secondary outcome)	Significantly fewer hip fractures for screening (2.6%) compared to usual care (3.5%) (HR 0.72 95%CI 0.59 to 0.89, p=0.002). A 28% relative reduction in hip fracture with screening	Intention-to- treat analysis: No significant difference between screening (3.1%) and usual care (3.1%) (SHR 1.002 95%CI 0.89 to 1.13, p=0.97) Pre-specified per-protocol analysis: No significant difference between screening (1.8%) and usual care (2.2%) (SHR 0.821 95%CI 0.67 to 1.01, p=0.06)

†††† Fracture, excluding the hands, feet, nose, skull or cervical vertebrae

‡‡‡‡ Hip, clinical vertebral, wrist or humerus fracture

§§§§ The Fine-Gray competing risk regression model was used. Sub-hazard ratios (SHR) were reported which consider the individual effect of a variable after accounting for other variables in the model. In this case death was counted as a competing risk and emigration as a censoring event

	SCOOP[12]	ROSE[11]
Mortality (secondary outcome)	No significant difference between screening (8.8%) and usual care (8.4%) (HR 1.05 95%CI 0.93 to 1.19, p=0.436)	Not reported
Anxiety levels (secondary outcome)	No significant difference between groups (p=0.515). Mean (standard deviation) scores were 10.5 (3.83) for low risk women; 10.6 (3.70) for high risk women and 10.4 (3.81) for usual care	Not reported
Quality of life (secondary outcome)	No significant difference between groups: • EQ-5D: screening 0.63 (0.33); usual care 0.63 (0.32); p=0.154 • SF-12 physical health: screening 38.3 (16.7); usual care 38.3 (16.6); p=0.237 • SF-12 mental health: screening 46.0 (18.3); usual care 46.3 (18.2); p=0.554	Not reported

In addition to the intention-to- treat and pre-specified per-protocol analysis reported in Table 5, Rubin et al (2018)[11] also reported a post-hoc analysis. This reported significant differences in major osteoporotic fracture, any fracture and hip fracture outcomes but this analysis only included a sub-group of women from the screening and usual care groups who had a FRAX score of ≥15% (ie were at moderate to high risk 10-year risk of major osteoporotic fracture) . Further details are provided in Appendix 3.

The SCOOP authors reported that, based on the absolute size of the decrease in hip fracture rate of 0.9%, 111 individuals would need to be screened to avert 1 hip fracture[12]. The ROSE authors reported that 1 hip fracture would have been prevented for approximately every 300 women screened and 1 major osteoporotic fracture for approximately every 150 women screened[11], using the per-protocol analysis. However, measures of impact will be dependent on the women who participate in a screening programme and the fracture incidence in the population.

The RCTs were assessed using the Cochrane Collaboration's tool for assessing risk of bias. Further details are provided in the Appendix 3 tables. There were no concerns regarding the randomisation process used. Both studies were potentially at high risk of contamination of the usual care group through raised awareness of osteoporosis in women and GPs through the existence of the study and national guidelines. For example, national or local guidelines that were in place or introduced during the trials include assessing fracture risk, offering DEXA scans and treating women as part of standard clinical practice. These factors may have diluted differences in outcomes between the trial arms. Neither trial used blinding for the participants or GPs but this would not have been possible given the nature of the trial. Blinding could potentially have been used in the assessment of outcomes, however the risk of bias was low due to the nature of the outcomes which were taken from existing data sets. The SCOOP trial included some self-reported quality of life outcomes which were more at risk of bias from lack of blinding and study attrition. However, the proportion of self-reported data received was fairly high (over 85% for both groups). The studies provided only limited details on the advice or treatment received by women in both study groups, only reporting number of prescriptions for anti-osteoporotic medication. Neither study provided any details of other interventions received by patients eg advice or lifestyle interventions. Both studies reported pre-specified outcomes and included intention-to-treat analysis.

Issues noted by the study authors with regards to the study sample included a lower mortality rate than was expected in the SCOOP trial (9% compared to an expected rate of 19%). The SCOOP trial also reported that the proportion of women screened who were classified as high risk was lower than expected for post-menopausal women (14% vs. 20-40%). However rates of fracture observed were higher than predicted before study commencement. The ROSE trial authors noted that women in the screening group who received a DEXA scan were younger and less likely to smoke than the control group and therefore may have been at lower risk of fracture.

Both of these trials had reasonably large sample sizes and assessed the effectiveness of population or community screening compared to usual care and both used risk assessment and DEXA in the selection of women for intervention. Both trials had a 5 year follow-up period, which may be considered relatively short in the context of the 10-year time frame used to assess fracture risk in the FRAX risk assessment tool.

There were differences in the approach used in these trials, eg the SCOOP trial testing algorithm assessed women for risk of hip fracture using FRAX and the ROSE trial assessed women for risk of major osteoporotic fracture using FRAX. For the SCOOP trial, this resulted in a discrepancy between the fracture risk screened for (hip fracture) and the primary outcome of the study (any osteoporotic fracture). The study authors suggest that this may provide a potential explanation for the study results in which a significant difference for hip fractures, but not osteoporosis-related fracture was observed.

The age range also varied between the studies with the SCOOP trial including women aged 70 to 85 years and the ROSE trial women aged 65 to 80. The significant difference in hip fractures reported by the SCOOP trial but not in the ROSE trial's planned analyses may reflect the fact that the SCOOP trial used risk of hip fracture as the screening test. The different approaches taken in the 2 trials in the risk assessed (eg hip fracture or any fracture) and the slightly different age groups targeted also raises questions about the risk to be assessed for a population screening programme (also discussed in relation to question 1).

The SCOOP trial authors reported that women who chose to participate reported better education, higher socioeconomic status and more frequent history of previous fracture or parental hip fracture compared to women who declined to participate. The ROSE trial authors reported that women who chose not to participate in the study (non-responders to the questionnaire) were more likely to

be older, have a lower personal income and education level, live alone and have co-morbidities. The study authors found that the majority of major osteoporotic fractures (56%) and hip fractures (65%) occurred in women who did not return a questionnaire, suggesting that the trial did not capture all women at high risk of fracture.

Summary of Findings Relevant to Criteria 11,12 and 13: Criteria not met*****

The previous 2013 UK NSC review did not identify any relevant RCTs of screening.

The results of 2 recent and relevant RCTs were available for the 2018 review, both of which compared population or community screening to usual care, used tests recommended for use in the UK and were conducted in either a UK population or a country analogous to the UK. Both RCTs found that systematic screening strategies using FRAX and DEXA in post-menopausal women do not significantly reduce osteoporotic fracture. A pre-specified secondary outcome in 1 of the trials suggested that there may be an impact for screening on hip fracture but conclusions based on secondary outcomes should be treated with caution.

The results of these trials are insufficient to conclude that the UK NSC criteria are met. However, the reduction in hip fractures may warrant further investigation.

These criteria are not met.

***** **Met** -for example, this should be applied in circumstances in which there is a sufficient volume of evidence of sufficient quality to judge an outcome or effect which is unlikely to be changed by further research or systematic review.
Not Met - for example, this should be applied in circumstances where there is insufficient evidence to clearly judge an outcome or effect or where there is sufficient evidence of poor performance.
Uncertain -for example, this should be applied in circumstances in which the constraints of an evidence summary prevent a reliable answer to the question. An example of this may be when the need for a systematic review and meta-analysis is identified by the rapid review.

Criterion 14

14. The opportunity cost of the screening programme (including testing, diagnosis and treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (ie value for money). Assessment against this criterion should have regard to evidence from cost benefit and/or cost-effectiveness analyses and have regard to the effective use of available resource.

Question 4 – Have UK evaluations demonstrated that screening for osteoporosis is cost-effective?

This question was not considered by the previous 2013 UK NSC review.

Eligibility for inclusion in the review

The inclusion criteria for this question are summarised briefly below:

- population – postmenopausal women
- intervention – screening – any combination of fracture risk assessment + DEXA
- comparator – no screening. Standard care
- outcomes – cost-effectiveness measured by QALYs
- study design – cost-effectiveness analyses; technology assessments; systematic reviews and meta-analyses; modelling studies. Only studies conducted in the UK which factor in recent RCT evidence
- date and language – studies published in English after 1st January 2011

Description of the evidence

Database searches yielded 780 results, of which 7 were judged to be relevant to this question and 1 abstract met the criteria for full text review. After review of the full text this study was included.

Discussion of findings

A study-level summary of data extracted from each included publication is presented in the summary and appraisal of individual studies in Appendix 3. In Appendix 3 publications are stratified by question.

Turner et al (2018)[13] performed an economic evaluation based on data from the UK SCOOP trial[12]. Further details of this trial are provided in the response to question 3. In their evaluation Turner et al included costs for the identification of women, the screening process and results, medications, and costs associated with GP and hospital attendances. It did not include any costs associated with oversight of the screening process. Quality of life outcomes used in the modelling were based on the EQ-5D results taken from the RCT and a 5 year time horizon was used.

In the primary analysis, the quality adjusted life year (QALY) gained by screening was small (0.0237) and not significantly different compared to usual care. However, the study authors[13] noted that because this is a screening intervention, the majority of participants in the screening arm received no change in the health care and would therefore not be expected to generate a large QALY gain. The study authors also noted that the baseline EQ-5D scores were lower in the intervention than usual care group and suggest that this would bias the QALY estimates in favour on the usual care group. In the secondary analysis, significant reductions in cost per fracture prevented were found for screening in the base case analysis, but not the sensitivity analysis. The wide confidence intervals around the estimates should be noted, which reduce confidence in the results.

The results are summarised in Table 6 below.

Table 6. Summary of results (Turner et al 2018[13])

		Base case analysis (full and imputed data set†††††)	Sensitivity analysis (CCA data set‡‡‡‡‡)
Primary analysis	Discounted QALY§§§§§	No significant difference between screening and usual care 0.008 (95%CI -0.028 to 0.044)	No significant difference between screening and usual care -0.005 (95%CI -0.051 to 0.040)
Primary analysis	Incremental QALY******	No significant difference in effect between screening and usual care Incremental effect: 0.0237 (95%CI -0.003 to 0.051); Incremental cost: £66 (95%CI -21.7 to 153); ICER††††††: £2,772	No significant difference in effect between screening and usual care Incremental effect: 0.0214 (95%CI -0.011 to 0.054); Incremental cost: £99 (95%CI 3 to 196); ICER: £4,646
Secondary analysis	Osteoporotic fracture prevented	Significant difference in effect between screening and usual care Incremental effect: 0.0146 (95%CI 0.0002 to 0.029); Incremental cost: £65 (95% -23.7 to 154.5); ICER: £4,478	No significant difference in effect between screening and usual care Incremental effect: 0.0094 (95%CI -0.007 to 0.026) Incremental cost: £99 (95%CI 3.2 to 195.5) ICER: £10,564
Secondary analysis	Hip fracture prevented	Significant difference in effect between screening and usual care Incremental effect: 0.0085 (95%CI 0.003 to 0.014); Incremental cost: £65 (95% -23.4 to 154.1); ICER: £7,694	No significant difference in effect between screening and usual care Incremental effect: 0.0045 (95%CI -0.002 to 0.011) Incremental cost: £99 (95%CI 3.4 to 195.2) ICER: £22,067

CCA – complete case analysis; CI – confidence interval; ICER - incremental cost effectiveness ratio; QALY – quality adjusted life year

††††† A full quality of life data set for a patient required 7 EQ-5D questionnaires over the 5 year follow-up period. Imputation was used where participants were missing data or where questionnaires had not been returned

‡‡‡‡‡ A complete case analysis (CCA) data set was used to evaluate the effect of only using cases where QALY could be estimated without multiple imputation. The CCA data set did include cases where 1 EQ-5D was missing. A hot-decking method was used where data for the missing question was imputed by comparing the 4 completed responses with patients with complete data who had the same pattern of responses to those 4 responses

§§§§§ Discounted QALY scores (unadjusted) were imputed using the following variables: baseline EQ-5D, age at randomisation, days alive, time without osteoporotic fracture and time without hip fracture

****** These results adjust for differences in baseline age and EQ-5D

†††††† Additional cost per QALY compared with usual care

The probabilities that screening would be cost-effective at a threshold of £20,000 per QALY in the base case and sensitivity analysis were 93% and 83%. However, this should be considered within the context of the lack of a significant difference between screening and usual care in the primary outcome of osteoporotic fracture in the source RCT.

The evaluation was assessed using the CASP checklist for economic evaluations. There were no concerns in the design or type of data sources used. The evaluation used a UK context and discounting was included. Sensitivity analysis was performed. The proportion of missing data was high, with a full data set available for approximately 55% of RCT participants. The 5 year time horizon reflects the 5 year follow-up timeframe of the RCT. However, this is shorter than the 10-year timeframe used to assess risk of hip fracture in the FRAX tool used in the RCT.

Summary of Findings Relevant to Criterion 14: Criterion uncertain‡‡‡‡‡‡

This question was not considered by the previous 2013 UK NSC review.

One study has assessed the cost-effectiveness of screening based on data from a UK RCT. There was a lack of significant difference between screening and usual care in the QALY estimates. The secondary analysis found significant reductions in cost per fracture prevented with screening in the base case analysis, but not the sensitivity analysis. However, the effect sizes were small and the confidence intervals wide reducing confidence in the results.

The modelling performed included some, but not all of the costs associated with a population screening programme and included within this criterion. The data provided by the screening trials discussed in question 3 could be used to form the basis of a wider evaluation of the cost-effectiveness of a population screening programme. However, the value of this analysis needs to be considered within the context of the lack of a significant difference between

‡‡‡‡‡‡ **Met** -for example, this should be applied in circumstances in which there is a sufficient volume of evidence of sufficient quality to judge an outcome or effect which is unlikely to be changed by further research or systematic review.
Not Met - for example, this should be applied in circumstances where there is insufficient evidence to clearly judge an outcome or effect or where there is sufficient evidence of poor performance.
Uncertain -for example, this should be applied in circumstances in which the constraints of an evidence summary prevent a reliable answer to the question. An example of this may be when the need for a systematic review and meta-analysis is identified by the rapid review.

screening and usual care in the primary outcome of both trials. At present this criterion is uncertain.

Review summary

Conclusions and implications for policy

The aim of a national screening programme targeting postmenopausal women would be to prevent osteoporotic fracture. Two RCTs have investigated the effectiveness of screening compared to usual care, one conducted in the UK and the other in Denmark. Neither RCT found an advantage for screening over usual care in the prevention of osteoporotic fracture as a whole although a reduction in hip fracture with screening was reported as a secondary outcome in the UK RCT. In the Danish trial prevention of osteoporotic, all cause and hip fractures were reported in a post hoc analysis of a sub-group of patients which should be considered exploratory.

On the current evidence base, a national screening programme cannot be recommended. Important areas of uncertainty remain:

- there are different potential approaches to screening for osteoporosis or risk of different types of fracture with varying test performance and a lack of studies in the specific population of interest
- concerns remain about the limited impact of pharmacological intervention on non-vertebral fractures and uncertainty about whether the evidence on the effectiveness of intervention is generalisable to a screened population
- there is an absence of studies demonstrating that screening has an impact on fracture-related morbidity or mortality
- there is a lack of evidence demonstrating an advantage to population screening over usual care from randomised controlled trials. However, the reduction in hip fracture observed as a secondary outcome in 1 RCT may warrant further research. Hip fractures can have long term health consequences, increase the risk of other (physical) problems, and so have negative impacts on quality of life.
- the opportunity cost of a full population screening programme is uncertain.

It is possible that increased awareness of the benefits of identifying and treating osteoporosis and the implementation of national guidelines within usual care

limits the ability of studies to demonstrate an advantage to population screening. The direction of future research and reviews may benefit from focusing on a specific target condition (eg osteoporosis or risk of any, or a specific, osteoporotic fracture) in order to increase the volume of comparable evidence to further inform decisions about population screening for osteoporosis.

Further consideration and agreement concerning the focus of a population screening programme (eg osteoporosis or risk of any of specific fragility fractures), would help direct future research and reviews. There is also uncertainty about how much additional value would be gained by population screening over the current usual care.

Limitations

A limitation for this review is the lack of studies that meet all of the inclusion criteria for this review for some of the key questions.

This rapid review process was conducted over a condensed period of time (approximately 12 weeks). Searching was limited to 3 bibliographic databases and did not include grey literature sources. The review was guided by a protocol developed a priori. The literature search and first appraisal of search results were undertaken by 1 information scientist, and further appraisal and study selection by 1 reviewer. Any queries at both stages were resolved through discussion with a second reviewer. Studies not available in the English language, abstracts and poster presentations, were not included. Studies that were not published in peer-reviewed journals were not reviewed.

Appendix 1 — Search strategy

Electronic databases

The search strategy included searches of the databases shown in Table 7.

Table 7. Summary of electronic database searches and dates

Database	Platform	Searched on date	Date range of search
MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print	Ovid SP	13th February 2017 6th September 2018	January 2011 to February 2017 February 2017 to September 2018
Embase	Ovid SP	13th February 2017 6th September 2018	January 2011 to February 2017 February 2017 to September 2018
The Cochrane Library, including: - **Cochrane Database of Systematic Reviews (CDSR)** - **Cochrane Central Register of Controlled Trials (CENTRAL)** - **Database of Abstracts of Reviews of Effects (DARE)**	Wiley Online	13th February 2017 6th September 2018	January 2011 to February 2017 February 2017 to September 2018

Search Terms

Search terms included combinations of free text and subject headings. The search was run twice in February 2017 and September 2018. Search terms for MEDLINE are shown in Table 8. Similar searches were used for Embase. Search terms for the Cochrane Library databases are shown in Table 9.

Table 8. Search strategy for MEDLINE

Term Group	Search terms
Question 1	
1	Women/
2	(wom?n or female$).tw.
3	1 or 2
4	(ABONE or age body size no estrogen).tw.

5	(DOEScore or Dubbo Osteoporosis Epidemiology Study).tw.
6	(MORES or Multiple Outcomes of Raloxifene Study).tw.
7	(NOF Guideline or National Osteoporosis Foundation).tw.
8	(OPERA or Osteoporosis Prescreening Risk Assessment).tw.
9	(ORAI or Osteoporosis Risk Assessment).tw.
10	(OSIRIS or Osteoporosis Index of Risk).tw.
11	(OST or Osteoporosis Self-assessment Screening Tool).tw.
12	(SCORE or Simple Calculated Osteoporosis Risk Estimation Study).tw.
13	(SOF or Study of Osteoporotic Fractures Study).tw.
14	(SOFSURF or Study of Osteoporosis Fractures Study Utilizing Risk Factors).tw.
15	(EPESE or Established Population for Epidemiology Studies of the Elderly Study).tw.
16	Fracture index.tw.
17	FRAX.tw.
18	Garvan nomogram.tw.
19	Minimum data set.tw.
20	QFracture.tw.
21	(WHI or Women's Health Initiative).tw.
22	lifestyle questionnaire.tw.
23	((fracture or osteporosis or risk or bone mass density) adj assessment tool$).tw.
24	Risk Assessment/
25	4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24
26	(DEXA or DXA or dual energy x-ray absorpitometry).tw.
27	(QUS or quantitative ultrasonography).tw.
28	26 or 27
29	25 and 28
30	"sensitivity and specificity"/
31	(sensitiv$ or specific$).tw.
32	Predictive Value of Tests/
33	(PPV or positive predictive value$ or NPV or negative predictive value$).tw.
34	((False or true) adj (negative$ or positive$)).tw.
35	likelihood ratio$.tw.
36	(AUC or area under the curve).tw.
37	(test adj2 (accura$ or reliab$ or valid$)).tw.
38	30 or 31 or 32 or 33 or 34 or 35 or 36 or 37
39	3 and 29 and 38
40	limit 39 to yr="2017 -Current"
Question 2	
1	women/
2	(wom?n or female$).tw.
3	1 or 2
4	Primary Prevention/
5	Therapeutics/
6	(prevent$ or treat$ or therap$ or intervention$).ti.
7	4 or 5 or 6
8	Bone Density Conservation Agents/
9	Diphosphonates/
10	(diphosphonate$ or bisphosphonate$).tw.
11	Alendronate/
12	alendronate.tw.
13	Etidronic Acid/
14	etidronate.tw.
15	ibandronate.tw.
16	Risedronate Sodium/
17	risedronate.tw.

18	zoledronate.tw.
19	Denosumab/
20	denosumab.tw.
21	Raloxifene Hydrochloride/
22	raloxifene.tw.
23	strontium.tw.
24	Teriparatide/
25	parathyroid hormone peptide$.tw.
26	Vitamin D/
27	vitamin D.tw.
28	Calcium, Dietary/
29	(calcium adj2 supplement$).tw.
30	exp Exercise/
31	(exercise or physical activity).tw.
32	(fall$ adj2 prevent$).tw.
33	((alcohol or smoking or cigarette) adj4 (reduc$ or cessation or stop$)).tw.
34	(lifestyle adj (modification$ or intervention$)).tw.
35	8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34
36	Osteoporotic Fractures/
37	((osteoporotic or clinical) adj fracture$).tw.
38	((vertebral or non vertebral or hip) adj fracture$).tw.
39	((reduc$ or prevent$) adj4 fracture$).tw.
40	36 or 37 or 38 or 39
41	3 and 7 and 35 and 40
42	limit 41 to yr="2017 -Current"
Question 3	
1	women/
2	(wom?n or female$).tw.
3	1 or 2
4	Primary Prevention/
5	Therapeutics/
6	(prevent$ or treat$ or therap$ or intervention$).ti.
7	Bone Density Conservation Agents/
8	Diphosphonates/
9	(diphosphonate$ or bisphosphonate$).tw.
10	Alendronate/
11	alendronate.tw.
12	Etidronic Acid/
13	etidronate.tw.
14	ibandronate.tw.
15	Risedronate Sodium/
16	risedronate.tw.
17	zoledronate.tw.
18	Denosumab/
19	denosumab.tw.
20	Raloxifene Hydrochloride/
21	raloxifene.tw.
22	strontium.tw.
23	Teriparatide/
24	parathyroid hormone peptide$.tw.
25	Vitamin D/
26	vitamin D.tw.
27	Calcium, Dietary/
28	(calcium adj2 supplement$).tw.
29	exp Exercise/

30	(exercise or physical activity).tw.
31	(fall$ adj2 prevent$).tw.
32	((alcohol or smoking or cigarette) adj4 (reduc$ or cessation or stop$)).tw.
33	(lifestyle adj (modification$ or intervention$)).tw.
34	4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33
35	Mass Screening/
36	(screen$3 or detect$3 or test or tests or testing).tw.
37	35 or 36
38	Osteoporosis/
39	Osteoporosis, Postmenopausal/
40	(osteoporosis or osteopenia).tw.
41	38 or 39 or 40
42	3 and 34 and 37 and 41
43	limit 42 to (english language and yr="2017 -Current")
44	exp animals/ not humans.sh.
45	43 not 44
Question 4	
1	women/
2	(wom?n or female$).tw.
3	1 or 2
4	Mass Screening/
5	(screen$3 or detect$3 or test or tests or testing).tw.
6	4 or 5
7	Osteoporosis/
8	Osteoporosis, Postmenopausal/
9	(osteoporosis or osteopenia).tw.
10	Osteoporotic Fractures/
11	((osteoporotic or clinical) adj fracture$).tw.
12	((vertebral or non vertebral or hip) adj fracture$).tw.
13	((reduc$ or prevent$) adj4 fracture$).tw.
14	7 or 8 or 9 or 10 or 11 or 12 or 13
15	3 and 6 and 14
16	(ABONE or age body size no estrogen).tw.
17	(DOEScore or Dubbo Osteoporosis Epidemiology Study).tw.
18	(MORES or Multiple Outcomes of Raloxifene Study).tw.
19	(NOF Guideline or National Osteoporosis Foundation).tw.
20	(OPERA or Osteoporosis Prescreening Risk Assessment).tw.
21	(ORAI or Osteoporosis Risk Assessment).tw.
22	(OSIRIS or Osteoporosis Index of Risk).tw.
23	(OST or Osteoporosis Self-assessment Screening Tool).tw.
24	(SCORE or Simple Calculated Osteoporosis Risk Estimation Study).tw.
25	(SOF or Study of Osteoporotic Fractures Study).tw.
26	(SOFSURF or Study of Osteoporosis Fractures Study Utilizing Risk Factors).tw.
27	(EPESE or Established Population for Epidemiology Studies of the Elderly Study).tw.
28	Fracture index.tw.
29	FRAX.tw.
30	Garvan nomogram.tw.
31	Minimum data set.tw.
32	QFracture.tw.
33	(WHI or Women's Health Initiative).tw.
34	lifestyle questionnaire.tw.
35	((fracture or osteporosis or risk or bone mass density) adj assessment tool$).tw.
36	Risk Assessment/

37	16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36
38	(DEXA or DXA or dual energy x-ray absorpitometry).tw.
39	(QUS or quantitative ultrasonography).tw.
40	38 or 39
41	37 and 40
42	3 and 41
43	15 or 42
44	limit 43 to ("economics (best balance of sensitivity and specificity)" or "costs (best balance of sensitivity and specificity)")
45	limit 44 to (english language and yr="2011 -Current")

Table 9. Search strategy for the Cochrane Library Databases

#	Search terms
#1	MeSH descriptor: [Women] explode all trees
#2	((woman or women or femal*)):ti,ab,kw
#3	#1 or #2
#4	(ABONE or "age body size no estrogen"):ti,ab,kw
#5	(DOEScore or "Dubbo Osteoporosis Epidemiology Study"):ti,ab,kw
#6	(MORES or "Multiple Outcomes of Raloxifene Study"):ti,ab,kw
#7	(("NOF Guideline" or "National Osteoporosis Foundation")):ti,ab,kw
#8	((OPERA or "Osteoporosis Prescreening Risk Assessment")):ti,ab,kw
#9	((OPERA or "Osteoporosis Prescreening Risk Assessment")):ti,ab,kw
#10	((OSIRIS or "Osteoporosis Index of Risk")):ti,ab,kw
#11	((OST or "Osteoporosis Self-assessment Screening Tool")):ti,ab,kw
#12	((SOF or "Study of Osteoporotic Fractures Study")):ti,ab,kw
#13	(SCORE or "Simple Calculated Osteoporosis Risk Estimation Study"):ti,ab,kw
#14	((SOFSURF or "Study of Osteoporosis Fractures Study Utilizing Risk Factors")):ti,ab,kw
#15	((EPESE or "Established Population for Epidemiology Studies of the Elderly Study")):ti,ab,kw
#16	("Fracture index"):ti,ab,kw
#17	(FRAX):ti,ab,kw
#18	("Garvan nomogram"):ti,ab,kw
#19	("Minimum data set"):ti,ab,kw
#20	(QFracture):ti,ab,kw
#21	((WHI or "Women's Health Initiative")):ti,ab,kw
#22	("lifestyle questionnaire*"):ti,ab,kw
#23	(("fracture assessment tool*" or "osteporosis assessment tool*" or "risk assessment tool*" or "bone mass density assessment tool*")):ti,ab,kw
#24	MeSH descriptor: [Risk Assessment] explode all trees
#25	#4 o r#5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24
#26	(DEXA or DXA or dual energy x-ray absorpitometry):ti,ab,kw
#27	((QUS or quantitative ultrasonography)):ti,ab,kw
#28	#26 or #27
#29	#25 and #28
#30	#3 and #29
#31	MeSH descriptor: [Osteoporotic Fractures] explode all trees
#32	vertebral fracture* or "non vertebral fracture*" or "hip fracture*"
#33	("osteoporotic fracture" or "clinical fracture"):ti,ab,kw
#34	((reduct* and fract*)):ti,ab,kw
#35	((prevent* and fract*)):ti,ab,kw
#36	#31 or #32 or #33 or #34 or #35

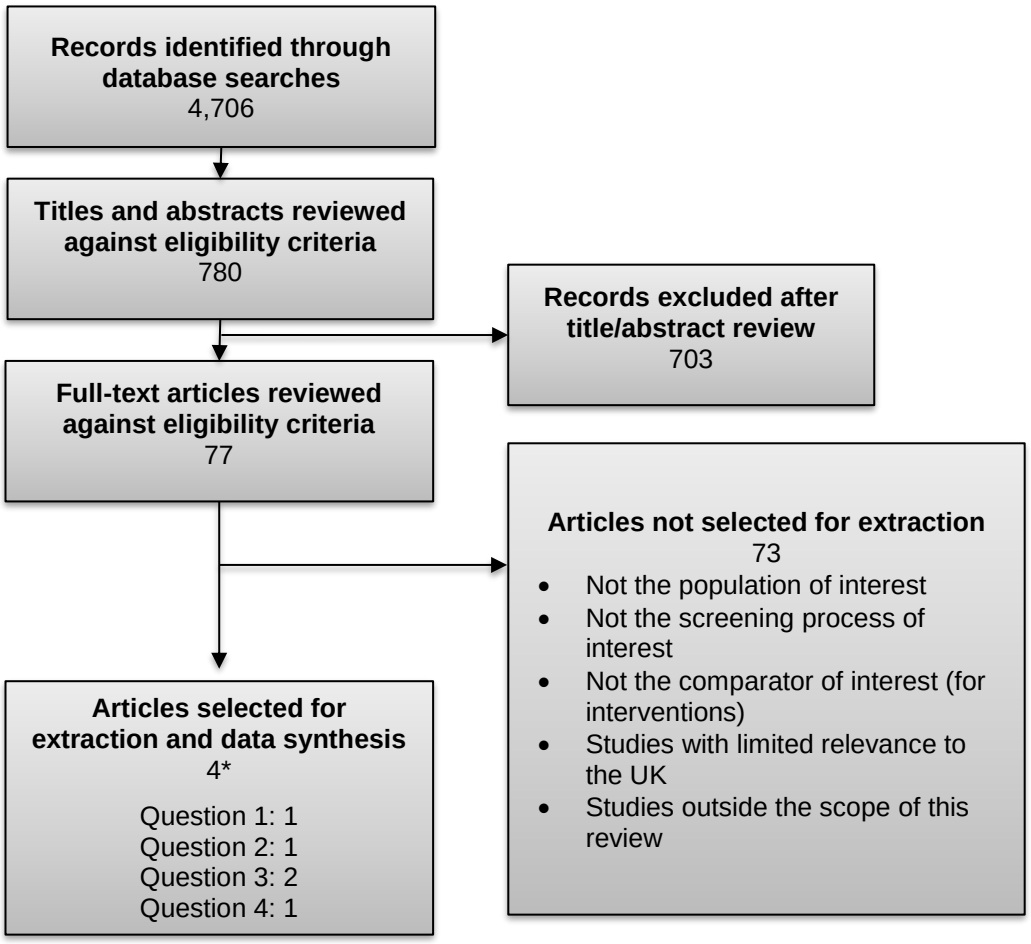
#37	MeSH descriptor: [Primary Prevention] explode all trees
#38	MeSH descriptor: [Therapeutics] explode all trees
#39	(prevent* or treat* or therap* or intervention*):ti
#40	MeSH descriptor: [Risk Reduction Behavior] explode all trees
#41	("lifestyle intervention*"):ti
#42	#37 or #38 or #39 or #40 or #41
#43	#3 and #36 and #42
#44	MeSH descriptor: [Mass Screening] explode all trees
#45	(screen* or detect* or test or tests or testing):ti,ab,kw
#46	#44 or #45
#47	MeSH descriptor: [Osteoporosis] explode all trees
#48	(osteoporosis or osteopenia):ti,ab,kw
#49	#47 or #48
#50	#3 and #46 and #49

Appendix 2 — Included and excluded studies

PRISMA flowchart

Figure 1 summarises the volume of publications included and excluded at each stage of the review. Seventy-seven publications were ultimately judged to be relevant to 1 or more review questions and were considered for extraction. Publications that were included or excluded after the review of full-text articles are detailed below.

Figure 1. Summary of publications included and excluded at each stage of the review



* The same systematic review was selected for both questions 1 and 2

Publications included after review of full-text articles

The 4 publications included after review of full-texts are summarised in Table 10.

Studies were prioritised for extraction and data synthesis. It was planned *a priori* that the following approach would be taken to prioritise studies for extraction:

1. Systematic reviews and meta-analyses would be considered the highest quality of evidence if any were found.

Table 10. Summary of publications included after review of full-text articles, and the question(s) each publication was identified as being relevant to

Study	The test	The intervention	The screening programme	Comments
Viswanathan et al (2018)[8]	X	X		
Shepstone et al (2018)[12]			X	
Rubin et al (2018)[11]			X	
Turner et al (2018)[13]			X	

Appendix 3 — Summary and appraisal of individual studies

Data Extraction and quality assessment for studies relevant to criteria 4 and 5

Key question 1: What is the accuracy of screening tests for osteoporosis?

Table 11. Viswanathan et al (2018)[8] (screening tests)

Publication	Viswanathan M. Reddy S. Berkman N. et al. Screening to prevent osteoporotic fractures: updated evidence report and systematic review for the U.S. Preventative Services Task Force. JAMA 2018, 319(24): 2532-2551
Study details	Systematic review
Study objectives	To assess screening to prevent osteoporotic fractures for 5 key questions. This included a question on the accuracy and reliability of screening approaches to identify adults at increased risk of osteoporotic fracture
Inclusions	Studies published between November 2009 and March 2018 where the majority of participants were community-dwelling adults with no known low-trauma fractures or metabolic bone disease
Exclusions	See inclusions
Population	The population for the review included adults (men and women) over 40 years old. Findings relating to women have been extracted for this review
Intervention	Studies on risk assessment tools and bone measurement testing, alone or in combination. Only results for tools that combine fracture risk assessment and BMD measurement are reproduced
Comparator	Not specified
Outcomes	**Fracture Risk Assessment Tool (FRAX) 10-year risk** (includes age, sex, weight, height, previous fracture, parental hip fracture, current smoking, glucocorticoid steroid use, rheumatoid arthritis, secondary osteoporosis, alcohol use, hip BMD (optional)) • For predicting major osteoporotic fracture including measurement of BMD: Pooled AUC 0.70 (95%CI 0.68 to 0.71; I^2=92.1%) (12 studies, n=62,054) • For predicting major osteoporotic fracture without measurement of BMD: Pooled AUC 0.66 (95%CI 0.63 to 0.69; I^2=99.2%) (17 studies, n=158,897) • For predicting hip fracture including measurement of BMD: Pooled AUC 0.79 (95%CI 0.76 to 0.81; I^2=99.1%) (10 studies, n=161,984) • For predicting hip fracture without measurement of BMD: Pooled AUC 0.76 (95%CI 0.72 to 0.81; I^2=99.8%) (12 studies, n=190,795) **Garvan nomogram/Fracture Risk Calculator 10-year risk** (includes age, sex, weight, previous non-traumatic fracture since age 50 years, fall within past 12 months, hip BMD (optional))

- For predicting major osteoporotic fracture including measurement of BMD: Pooled AUC 0.68 (95%CI 0.64 to 0.71; I^2=84.8%) (3 studies, n=6,174)
- For predicting major osteoporotic fracture without measurement of BMD: AUC 0.66 (95%CI 0.61 to 0.72) (1 study, n=600)
- For predicting any osteoporotic fracture including measurement of BMD: AUC 0.69 (95%CI not reported) (1 study, n=506)
- For predicting any osteoporotic fracture without measurement of BMD: AUC 0.65 (95%CI not reported) (1 study, n=506)
- For predicting hip fracture including measurement of BMD: Pooled AUC 0.72 (95%CI 0.66 to 0.79; I^2=97.3%) (4 studies, n=7,449)
- For predicting hip fracture without measurement of BMD: AUC 0.68 (95%CI not reported) (1 study, n=1,369)
- For predicting non-vertebral fracture including measurement of BMD: AUC 0.62 (95%CI not reported) (1 study, n=1,646)
- For predicting non-vertebral fracture without measurement of BMD: AUC 0.58 (95%CI not reported) (1 study, n=1,637)

Women's Health Initiative (WHI) 5 year risk (includes age, weight, height, self-reported health, previous fracture after age 55 years, race/ ethnicity, physical activity, smoking, parental hip fracture after age 40 years, diabetes treated with medications, glucocorticoid steroid use, hip BMD (optional))

- For predicting hip fracture including measurement of BMD: AUC 0.80 (95%CI 0.75 to 0.85) (1 study, n=10,750)
- For predicting hip fracture without measurement of BMD: AUC in 2 studies was 0.80 (95%CI 0.77 to 0.82) (n=10,750) and 0.82 (95%CI not reported) (n=13,353)

Fracture and Immobilization Score (FRISC) 1, 3, 5 or 10-year risk (includes age, weight, menopausal status, secondary osteoporosis, prior fracture, back pain, dementia, lumbar BMD)

- For predicting major osteoporotic fracture including measurement of BMD: AUC 0.73 (95%CI not reported) (1 study, n=400)
- For predicting long bone and vertebral fracture including measurement of BMD: AUC 0.69 (95%CI 0.64 to 0.73) (1 study, n=765)

Fracture Risk Score (FRISK) 5 or 10-year risk (includes age, weight, height, prior fracture, prior falls, lumbar and hip BMD (optional))

- For predicting major osteoporotic fracture including measurement of BMD: AUC 0.66 (95%CI 0.60 to 0.71) (1 study, n=600)
- For predicting major osteoporotic fracture without measurement of BMD: AUC 0.62 (95%CI 0.56 to 0.67) (1 study, n=600)

Fracture risk calculator (FRC) 10-year risk (includes age, sex, BMI, prior fracture, parental fracture, smoking, alcohol use, glucocorticoid steroid use, rheumatoid arthritis, secondary osteoporosis, race/ethnicity, BMD (optional))

- For predicting hip fracture including measurement of BMD: AUC 0.85 (95%CI 0.84 to 0.86) (1 study, n=94,489)
- For predicting hip fracture without measurement of BMD: AUC 0.83 (95%CI 0.82 to 0.84) (1 study, n=94,489)

Quality appraisal	The study was assessed using the CASP checklist for systematic reviews. There were no concerns about the conduct of the review.

	The source studies were a systematic review, judged to be of good quality by the study authors, and 13 individual studies judged to be of good or fair quality. No details of individual studies were provided but issues with the evidence base identified by the review authors included inconsistent reporting of which version of risk assessment tools were used and studies that had follow-up periods shorter than the time period covered by the risk assessment. Many tools had results from a single study. The heterogeneity between the studies used in the pooled analysis was very high. There was considerable variation in the sample sizes from a few hundred women to almost 200,000. One of the tools included (FRAX) is recommended for use in the UK by NICE.

Data Extraction and quality assessment for studies relevant to criteria 9 and 10

Key question 2: What is the effectiveness of interventions in reducing the risk of osteoporotic fracture?

Sub-question: Have any studies explored the effectiveness of interventions in reducing mortality?

Table 12. Viswanathan et al (2018)[8] (interventions)

Publication	Viswanathan M. Reddy S. Berkman N. et al. Screening to prevent osteoporotic fractures: updated evidence report and systematic review for the U.S. Preventative Services Task Force. JAMA 2018, 319(24): 2532-2551
Study details	Systematic review
Study objectives	To assess screening to prevent osteoporotic fractures for 5 key questions. This included questions on the effectiveness of pharmacotherapy for the reduction of fractures and related morbidity and mortality and the harms associated with pharmacotherapy
Inclusions	Studies published between November 2009 and March 2018 where the majority of participants had an increased fracture risk
Exclusions	See inclusions
Population	The population for the review included adults (men and women) over 40 years old. Findings relating to women have been extracted for this review
Intervention	Pharmacotherapy including bisphosphonates, oestrogen agonists/antagonists, oestrogen and/or progestine-based hormone therapy, parathyroid hormone and RANK ligand inhibitors (eg denosumab)
Comparator	Placebo
Outcomes	Bisphosphonates • Bisphosphonates were associated with fewer vertebral fractures (2.1%) compared to placebo (3.8%) (Relative Risk (RR) 0.57 95%CI 0.41 to 0.78, I^2=0%) (5 RCTs, n=5,433) • Bisphosphonates were associated with fewer non-vertebral fractures (8.9%) compared to placebo (10.6%) (RR 0.84 95%CI 0.76 to 0.92, I^2=0%) (8 RCTs, n=16,438)

- No significant difference in hip fractures for bisphosphonates (0.70%) compared to placebo (0.96%) (RR 0.70 95%CI 0.44 to 1.11, I^2=0%) (3 RCTs, n=8,988)§§§§§§
- No significant difference in serious adverse events between bisphosphonates and placebo (RR 0.98 95%CI 0.92 to 1.04, I^2=0%) (17 RCTs, n=11,745)
- No significant difference in discontinuation between bisphosphonates and placebo (RR 0.99 95%CI 0.91 to 1.07, I^2=0%) (20 RCTs, n=17,369)
- No significant difference in upper gastrointestinal events between bisphosphonates and placebo (RR 1.01 95%CI 0.98 to 1.05, I^2=0%) (13 RCTs, n=20,485)
- 2 studies in women reported no cases of osteonecrosis of the jaw

Raloxifene

- Raloxifene reduced radiographic vertebral fracture (7.5%) compared to placebo (12.5%) (RR 0.64 95%CI 0.53 to 0.76) (1 RCT, n=7,705)
- No significant difference between raloxifene and placebo for incidence of non-vertebral fracture (details not reported)
- No significant difference between raloxifene and placebo for incidence of hip fracture (details not reported)
- No significant difference in discontinuation due to adverse events between raloxifene and placebo (RR 1.12 95%CI 0.98 to 1.28, I^2=0%) (6 RCTs, n=6,438)
- Possible (but non-significant) association between raloxifene and deep vein thrombosis (RR 2.14 95%CI 0.99 to 4.66, I^2=0%) (3 RCTs, n=5,839)
- No significant association between raloxifene and coronary heart disease (HR 0.88 95%CI 0.56 to 1.40) or stroke (RR 0.69 95%CI 0.40 to 1.18) (number of studies not reported)

Oestrogen

- Women taking oestrogen had lower risk of osteoporotic fracture compared to placebo (HR 0.72 95%CI 0.64 to 0.80) (from a systematic review; no of studies not reported)
- Women taking oestrogen plus progestin had a lower risk of fracture compared to placebo (RR 0.80 95%CI 0.68 to 0.94) (from a systematic review; no of studies not reported)
- A systematic review found that women receiving oestrogen with or without progesterone had higher rates of gall bladder events, stroke, venous thromboembolism and urinary incontinence and higher risk of invasive breast cancer, coronary heart disease and probable dementia (no further details reported)

Denosumab

- Denosumab reduced vertebral fracture (2.3%) compared to placebo (7.2%) (RR 0.32 95%CI 0.26 to 0.41) (1 RCT, n=7,868)
- Denosumab reduced non-vertebral fracture (6.1%) compared to placebo (7.5%) (RR 0.80 95%CI 0.67 to 0.95) (1 RCT, n=7,868)
- Denosumab reduced hip fracture (0.7%) compared to placebo (1.1%) (RR 0.60 95%CI 0.37 to 0.97) (1 RCT, n=7,868)
- No significant difference in serious adverse events between denosumab and placebo (RR 1.12 95%CI 0.88 to 1.44, I^2=14%) (4 RCTs, n=8,663)

§§§§§§ The review authors reported that only 1 of the 3 studies was sufficiently powered to detect differences in hip fracture

	• No significant difference in discontinuation due to adverse events between denosumab and placebo (RR 1.14 95%CI 0.85 to 1.52, I^2=0%) (3 RCTs, n=8,451) Parathyroid hormone • Parathyroid hormone reduced new radiographic vertebral fracture (0.7%) compared to placebo (2.1%) in women without a prevalent fracture at baseline (RR 0.32 95%CI 0.14 to 0.75) (1 RCT, n=2,061) • No significant difference in new non-vertebral fracture between parathyroid hormone (5.6%) and placebo (5.8%) in women with and without prevalent fracture (RR 0.97 95%CI 0.71 to 1.33) (1 RCT, n=2,532) • Postmenopausal women had significantly higher discontinuation rates with parathyroid hormone (30.2%) compared to placebo (24.6%) (RR 1.22 95%CI 1.08 to 1.40) (1 study, n=2,532)
Quality appraisal	The study was assessed using the CASP checklist for systematic reviews. There were no concerns about the conduct of the review. The source data for the review was RCTs which the review authors graded as being of either fair or good quality. Some of the outcomes were based on single large trials, other outcomes were reported to be dominated by a single big study for each drug. The heterogeneity between the studies used in the pooled analysis was low. The sample sizes, where reported, ranged from approximately 2,000 to over 20,000. The confidence intervals around some of the outcomes were wide reducing confidence in the results. The review authors graded the evidence base as a whole as low to moderate for reducing fractures. The review only considered pharmacological interventions for osteoporosis. The inclusion criteria for the systematic review were broader than the scope of this review. It is not clear how applicable the results are to the screening population of interest.

Data Extraction and quality assessment for studies relevant to criteria 11, 12 and 13

Key question 3: Have RCTs demonstrated the clinical benefit of screening in reducing osteoporotic fractures in comparison to standard care?

Table 13. Shepstone et al (2018)[12]

Publication	Shepstone L. Lenaghan E. Cooper C. et al. Screening in the community to reduce fractures in older women (SCOOP): a randomised controlled trial. Lancet 2018, 391: 714-747
Study details	RCT
Study objectives	To assess the effectiveness of a FRAX-based community screening programme for UK women aged 70 to 85 years in reducing the incidence of fractures over a 5 year period
Inclusions	Women aged 70 to 85 years

Exclusions	Women receiving prescription anti-osteoporotic drugs (excluding vitamin D or calcium)
Population	12,495 women identified from primary care lists from 100 GP practice in 7 UK regions who had consented to participate in the RCT. No individual practice could provide more than 500 participants. 12 patients (6 from each group) were excluded after randomisation. 12,483 patients were included in the analysis.
Intervention	Community screening (n=6,233) • Before randomisation, all women self-completed a baseline questionnaire which captured FRAX risk factors • Women at high risk of hip fracture on FRAX (10-year risk above an age-dependent threshold) were offered DEXA scan and the results were used to re-calculate risk of hip fracture • Women and their GP were informed whether they were at low or high risk of fracture • Women at high risk were advised to make an appointment with their GP to discuss treatment options
Comparator	Usual care (n=6,250) (usual care not defined) • Before randomisation, all women self-completed a baseline questionnaire which captured FRAX risk factors • Risk was not calculated for women in the usual care group until the end of the trial, for comparative purposes • Women and GPs were advised of the patient's participation in the study but received no other information

Outcomes	NHS data sets, primary care records and self-report. Only verified fractures were included. Patients were followed-up for 5 years. The analysis was intention-to-treat.
	
	Screening group:
	• 3,064 women (49%) were classed as high risk of hip fracture on FRAX and were offered a DEXA scan
	• 247 women (8%) did not have a DEXA scan (due to declining the invitation, unable to have a BMD measurement or died before scan)
	• 898 (14%) were classed as high risk after recalculation of their FRAX risk and were advised to discuss treatment options with their GP
	• Mean femoral neck T score was – 2.6 for high risk women
	
	Screening vs. usual care at 5 year follow-up
	Treatment outcomes:
	• 1,486 (24%) of the screening group received ≥1 prescription for anti-osteoporotic medication during the study compared to 982 (16%) of the control group (no significance test reported)
	Fracture outcomes:
	• 1,975 osteoporosis-related fractures were observed in 1,657 individuals (13% of the study population). This included:
	  • Distal forearm fracture: 638 fractures in 614 individuals
	  • Hip fracture: 392 fractures in 382 individuals
	Primary outcome:
	• No significant difference in osteoporosis-related fracture between screening (12.9%) and usual care (13.6%) (HR 0.94 95%CI 0.85 to 1.03, p=0.178)
	Secondary outcomes:
	• No significant difference in any clinical fracture between screening (15.3%) and usual care (16.0%) (HR 0.94 95%CI 0.86 to 1.03, p=0.183)

	• Significantly fewer hip fractures for screening (2.6%) compared to usual care (3.5%) (HR 0.72 95%CI 0.59 to 0.89, p=0.002); a 28% relative reduction in hip fractures with screening
	• No significant difference in mortality between screening (8.8%) and usual care (8.4%) (HR 1.05 95%CI 0.93 to 1.19, p=0.436)
	• No significant difference in anxiety levels between screening and usual care (p=0.515). Mean (SD) scores on the Short Form State-Trait Anxiety Inventory******* at 5 year follow-up were 10.5 (3.83) for low risk women; 10.6 (3.70) for high risk women and 10.4 (3.81) for usual care
	• No significant difference in quality of life between screening and usual care on EQ-5D (p=0.154), SF-12 physical health (p=0.237) or SF-12 mental health (p=0.554)
	  • Mean (SD) scores on the EQ-5D at 5 year follow-up were 0.63 (0.33) for screening and 0.63 (0.32) for usual care
	  • Mean (SD) scores on the SF-12 physical health at 5 year follow-up were 38.3 (16.7) for screening and 38.3 (16.6) for usual care
	  • Mean (SD) scores on the SF-12 mental health at 5 year follow-up were 46.0 (18.3) for screening and 46.3 (18.2) for usual care
	
	The study authors calculated that, based on the absolute size of the decrease in hip fracture rate of 0.9%, 111 individuals would need to be screened to avert 1 hip fracture.

Quality appraisal	The RCT was assessed using the Cochrane Collaboration's tool for assessing risk of bias. This was a large trial of a community-based screening programme. There were no selection bias concerns. The randomisation process used was block randomisation (block size 6) stratified by region and age group (70 to 74 years; 75 to 79 years; 80 to 85 years). There was low risk of performance and detection bias. Although no blinding was used in the study this would not have been possible for the participants or GPs. No blinding was reported for the assessment of outcomes but the risk of bias is low for non-subjective outcomes taken from existing data sets such as fracture. The study included some self-reported quality of life outcomes which were more at risk from lack of blinding and study attrition bias. However the proportion of self-reported data received remained fairly high during the study duration. At 5 year follow-up the proportion of self-reported data received was 85.6% in the screening group and 85.2% in the usual care group. The risk of reporting bias was low. The study authors performed an intention-to-treat analysis and reported pre-specified primary and secondary outcomes. <u>Other issues</u> Limited details were provided on the advice or treatment received by women in both study groups. The 5 year follow-up period may be considered relatively short in the context of the 10 year time frame used to assess fracture risk.

*******Scores on this scale range from 6 to 24, with lower scores indicating lower levels of anxiety

	The study authors reported that women who chose to participate in the trial reported better education, high socioeconomic status and more frequent history of previous fracture or parental hip fracture compared to women who declined to participate. The study authors noted that the study as a whole had a lower mortality rate than was expected for the eligible population (9% vs 19%). The proportion of women screened who were deemed at high risk was lower than expected for post-menopausal women (14% vs 20-40%). However, rates of fractures observed were higher than predicted in the power size calculation. It is possible that the existence of the study may have raised awareness about osteoporosis amongst GPs which may have affected monitoring and treatment of women in the usual care group. Recruitment to the study started in early 2008 with last follow-up in July 2014. The NICE clinical guideline on assessing the risk of fragility fracture in osteoporosis was published in August 2012 which may have influenced awareness and usual care. The FRAX assessment used as the screening test calculated 10-year risk of hip fracture rather than risk of any major osteoporotic fracture. This created a discrepancy between the fracture risk screened for and the study's primary outcome (any osteoporotic fracture). The women detected as high risk by the screening process were at high risk of hip fracture. The study authors suggest this may provide a potential reason why a significant difference in hip fractures but not any osteoporotic fractures was observed. Hip fractures represented 19.8% of the total osteoporosis-related fractures experienced by study participants. There was no difference between the screening and usual care groups on 2 quality of life measures.

Table 14. Rubin et al (2018)[11]

Publication	Rubin KH. Rothmann MJ. Holmberg T. et al. Effectiveness of a two-step population-based osteoporosis screening program using FRAX: the randomized Risk-stratified Osteoporosis Evaluation (ROSE) study. Osteoporosis International 2018, 29: 567-578
Study details	RCT
Study objectives	To investigate the effectiveness of a 2 step population-based osteoporosis screening programme
Inclusions	Women aged 65 to 80 years
Exclusions	N/a
Population	• Intention-to treat population: 34,229 women living in Southern Denmark who were sent a baseline questionnaire • Pre-specified per protocol population: 18,605 women who returned sufficient data to calculate FRAX and were not already receiving osteoporosis treatment • Post-hoc per protocol population: 5,009 women from the screening group who had a FRAX score ≥15% and received a DEXA scan and 7,026 women from the control group with a FRAX score ≥15
Intervention	Screening (n=9,279): • After randomisation, all women self-completed a baseline questionnaire which captured FRAX risk factors • Women at moderate to high risk of major osteoporotic fracture (10-year probability ≥15%) on FRAX were offered a DEXA scan

	• DEXA scans measured BMD. Osteoporosis was diagnosed if any of the measured sites had a T score of ≤-2.5 • Women and their GP were informed of the results. GPs received information on treatment recommendations based on national guidelines • Decisions on treatment were left to the patient and GP
Comparator	Usual care (n=9,326) • After randomisation, all women self-completed a baseline questionnaire which captured FRAX risk factors • Women were not informed about the result of their FRAX calculation
Outcomes	Data were taken from national registries including inpatient, outpatient and prescription data. Median follow-up was 5 years The primary analysis was proportion of women with major osteoporotic fracture, applied to the intention-to-treat population. Secondary analysis included 2 per-protocol analyses • 34,229 women sent a baseline questionnaire • 27,157 (79%) returned a baseline questionnaire (including blank questionnaires) • 20,905 (61%) questionnaires had sufficient information to calculate FRAX • 2,300 women were already receiving osteoporosis medication Screening group: • 7,056 (76%) had a FRAX ≥15% • 6,226 (67%) offered a DEXA scan • 5,009 (54%) received a DEXA scan • 1,236 (13%) recommended for treatment **Screening vs. usual care at 5 year follow-up** Treatment outcomes: • A significantly greater proportion of the screening group (23%) received osteoporosis medication during the study compared to the usual care group (18%) (p<0.001) Fracture outcomes: • 3,416 fractures were observed in the study population *Intention-to-treat analysis (all women sent a questionnaire):* Primary outcome: • No significant difference in major osteoporotic fracture††††††† between screening (9.9%) and usual care (10.0%) (SHR 0.986 95%CI 0.92 to 1.06, p=0.68) Secondary outcomes: • No significant difference in any fracture between screening (13.1%) and usual care (13.0%) (SHR‡‡‡‡‡‡‡ 1.004 95%CI 0.94 to 1.06, p=0.91)

††††††† Hip, clinical vertebral, wrist or humerus fracture

‡‡‡‡‡‡‡ The Fine-Gray competing risk regression model was used. Sub-hazard ratios were reported which consider the individual effect of a variable after accounting for other variables in the model. In this case death was counted as a competing risk and emigration as a censoring event

	• No significant difference in hip fracture between screening (3.1%) and usual care (3.1%) (SHR 1.002 95%CI 0.89 to 1.13, p=0.97) *Pre-specified per protocol analysis (women with a FRAX score and not already receiving osteoporosis medication)* Primary outcome: • No significant difference in major osteoporotic fracture**Error! Bookmark not defined.** between screening (7.8%) and usual care (8.4%) (SHR 0.914 95%CI 0.83 to 1.01, p=0.08) Secondary outcomes: • No significant difference in any fracture between screening (10.7%) and usual care (11.0%) (SHR 0.968 95%CI 0.89 to 1.06, p=0.47) • No significant difference in hip fracture between screening (1.8%) and usual care (2.2%) (SHR 0.821 95%CI 0.67 to 1.01, p=0.06) *Post-hoc per protocol analysis (DEXA scanned vs control FRAX ≥15%)* Primary outcome: • Significant reduction in major osteoporotic fracture for DEXA scanned (8.1%) compared to control (9.3%) (SHR 0.870 95%CI 0.77 to 0.99. p=0.03) Secondary outcomes: • Significant reduction in any fracture for DEXA scanned (10.8%) compared to control (12.1%) (SHR 0.892 95%CI 0.80 to 0.99, p=0.04) • Significant reduction in hip fracture for DEXA scanned (1.9%) compared to control (2.6%) (SHR 0.741 95%CI 0.58 to 0.95, p=0.02) The study authors used the pre-specified per protocol population to calculate that 1 hip fracture would have been prevented for approximately every 300 women screened and 1 major osteoporotic fracture for approximately every 150 women screened
Quality appraisal	The RCT was assessed using the Cochrane Collaboration's tool for assessing risk of bias. This was a large trial of a population-based screening programme. There were no selection bias concerns. Participants were randomised before their invitation to screening and stratified by area of residence and age (1 year age groups). There was low risk of performance, detection and study attrition bias. No blinding was used in the study but this would not have been possible for the participants or GPs. No blinding was reported for the assessment of outcomes but the objective nature of the outcomes (eg details of fracture taken from existing data sets) reduces the risk of detection or attrition bias. The proportion of participants who returned the baseline questionnaire was fairly high (79% and 80% in the screening and usual care groups), however this included questionnaires that were returned blank or had missing data. FRAX was calculated for 61% of randomised participants in both groups. DEXA was performed in 80% of women who were offered the scan. The study authors did not re-calculate FRAX score following DEXA. The risk of reporting bias was low. The study authors performed an intention-to-treat and pre-specified per-protocol analysis. A post-hoc per protocol analysis was also reported.

	Other issues Limited details were provided on the advice or treatment received by women in both study groups. The 5 year follow-up period may be considered relatively short in the context of the 10 year time frame used to assess fracture risk. It is possible that the completion of the questionnaire may have raised awareness about osteoporosis in the usual care group who may have sought advice from their GP and received intervention according to national guidance. It is also possible that the study may have raised awareness about the detection and treatment of osteoporosis amongst GPs. The study authors reported that women who chose not to participate in the study (non-responders to the questionnaire) were more likely to be older, have a lower personal income and education level, live alone and have co-morbidities. The study authors found that the majority of major osteoporotic fractures (56%) and hip fractures (65%) occurred in women who did not participate in the study by returning a questionnaire, suggesting that the study did not include all women at high risk of fracture. The study authors noted that women in the screening group who received a DEXA scan were younger and less likely to smoke than the control group and therefore may have been at lower risk of fracture. This introduces a potential source of bias into the post-hoc per protocol analysis.

Data Extraction and quality assessment for studies relevant to criterion 14

Key question 4: Have UK evaluations demonstrated that screening for osteoporosis is cost-effective?

Table 15. Turner et al 2018[13]

Publication	Turner DA. Khioe RFS. Shepstone L. et al. The Cost-Effectiveness of Screening in the Community to Reduce Osteoporotic Fractures in Older Women in the UK: Economic Evaluation of the SCOOP Study. Journal of Bone & Mineral Research 2018, 33(5): 845-51.
Study details	Economic evaluation conducted alongside a UK RCT of screening to prevent fractures
Study objectives	To assess the cost-effectiveness of a FRAX-based screening programme for older UK women
Inclusions	The economic evaluation included costs for: • Identification of eligible patients • Resource to administer screening questionnaires • Calculation of initial fracture risk algorithm • Notification of initial fracture risk, letters to participants and GPs • BMD assessment using DEXA scans • Calculation of final fracture risk • Clinical review of final fracture risk • Notification of final fracture risk result • GP consultations

	• Hospital admissions • Outpatient attendances • Procedure costs • A&E attendances • Anti-osteoporosis medications The cost for oversight of the screening process was reported as £0
Exclusions	The source RCT excluded women already on prescriptions for anti-osteoporosis medicines (except for vitamin D or calcium)
Population	The source RCT included 12,483 women aged 70 to 85 years. Women were followed-up for 5 years
Intervention	In the RCT screening arm, women with a 10-year hip fracture risk above an age-dependent threshold were recommended for treatment via their GP
Comparator	In the source RCT control arm, women received usual care, including referral for DEXA scan and anti-osteoporosis treatment if deemed clinically appropriate by their GP
Outcomes	Data for the economic evaluation were taken from NHS data sets, primary care records and self-report. Only verified fractures were included. Quality of life was assessed using the EQ-5D questionnaire. A 5 year time horizon was used. Costs were provided in pounds sterling for the year 2013/14 with discounting at a rate of 3.5%. The analysis used a within-trial, intention-to-treat basis from a UK NHS perspective. Full data§§§§§§§ was available for 6,881 patients (55% of the RCT population). This included 3,477 patients from the screening group (56%) and 3,404 patients from the control group (54%). Imputation was used where participants were missing ≥1 EQ-5D question or the questionnaire had not been returned. This was used to create the sample for base case analysis. A complete case analysis (CCA) set******** included 7,975 patients (64% of the RCT population) and was used in the sensitivity analysis. This provided a data set of cases where QALY could be estimated without multiple imputation. **Costs (per person)** • The average costs for the intervention were £104 Total, average discounted costs for intervention and fracture-related healthcare for the 5 year follow-up were reported for the full and CCA data sets: Full data sample: • Total costs for the intervention group: £968 • Total costs for the usual care group: £900 • Difference between intervention and usual care: £68 (95%CI -21 to 157) CCA sample: • Total costs for the intervention group: £833 • Total costs for the usual care group: £728

§§§§§§§ A full quality of life data set for a patient included 7 EQ-5D questionnaire returns over the 5 year follow-up period.

******** A 'hot-decking' method was applied to cases where patients had completed 4 of the 5 EQ-5D questions. Data for the missing question was imputed by comparing the 4 completed responses with patients with complete data who had the same pattern of responses to those 4 responses.

- Difference between intervention and usual care: £104 (95%CI 8 to 201)

The major component of costs was inpatient stay. The lower costs in the CCA sample reflects the lower proportion of fractures in the CCA sample compared to the full sample.

Primary analysis
Base case analysis (imputed sample)

- No significant difference in discounted QALY†††††††† between intervention (3.274) and control (3.266) 0.008 (95%CI -0.028 to 0.044)
- No significant difference in incremental QALY‡‡‡‡‡‡‡‡: 0.0237 (95%CI -0.003 to 0.051). Incremental cost £66 (95%CI -21.7 to 153). ICER £2,772

Sensitivity analysis (CCA data set)

- No significant difference in discounted QALY between intervention (3.368) and control (3.373) -0.005 (95%CI -0.051 to 0.040)
- No significant difference in incremental QALY: 0.0214 (95%CI -0.011 to 0.054). Incremental cost £99 (95%CI 3 to 196). ICER £4,646

Cost effectiveness acceptability curves showed that in the base case analysis there was a 93% probability that the intervention would be cost-effective at a threshold (used by NICE) of £20,000 per QALY. In the sensitivity analysis this was 83%.

Secondary analysis
Base case analysis (imputed sample)

- Significant difference in osteoporotic fracture prevented between intervention and control: incremental effect 0.0146 (95%CI 0.0002 to 0.029). Incremental cost £65 (95%CI -23.7 to 154.5). ICER £4,478
- Significant difference in hip fracture prevented between intervention and control: incremental effect 0.0085 (95%CI 0.003 to 0.014). Incremental cost £65 (95%CI -23.4 to 154.1). ICER £7,694

Sensitivity analysis (CCA data set)

- No significant difference in osteoporotic fracture prevented between intervention and control: incremental effect 0.0094 (95%CI -0.0073 to 0.026). Incremental cost £99 (95%CI 3.2 to 195.5). ICER £10,564
- No significant difference in osteoporotic fracture prevented between intervention and control: incremental effect 0.0045 (95%CI -0.002 to 0.011). Incremental cost £99 (95%CI 3.4 to 195.2). ICER £22,067

Cost effectiveness acceptability curves showed an 87% probability that intervention would be considered cost-effective if preventing a hip fracture was valued at £20,000. The figure for osteoporotic fracture was only displayed graphically and could not be accurately determined.

†††††††† Discounted QALY scores (unadjusted) were imputed using baseline EQ-5D, age at randomisation, days alive, time without osteoporotic fracture and time without hip fracture.
‡‡‡‡‡‡‡‡ Adjusted for differences in baseline age and EQ-5D

Quality appraisal	The evaluation was assessed using the CASP checklist for economic evaluations. There were no concerns in the design or type of data sources used. The evaluation included costs for the screening process, intervention and outcomes. The evaluation used a UK context and discounting was included. Sensitivity analysis was performed. The proportion of missing EQ-5D data was high with a full data set available for approximately 55% of the RCT participants. The study authors noted that the baseline EQ-5D scores were lower in the intervention than usual care group and suggest that this would bias the QALY estimates in favour of the control group. The confidence intervals around the estimates were very wide reducing confidence in the results. Differences between intervention and control were not statistically significant in the primary analysis. The study authors noted that the majority of participants in the screening arm of a screening trial receive no change in health care and therefore would not expect to generate a large QALY gain. The 5 year time horizon reflects the 5 year follow-up timeframe of the RCT. However, both of these are shorter than the 10-year timeframe used to assess risk of hip fracture in the FRAX tool used in the RCT.

Appendix 4 – UK NSC reporting checklist for evidence summaries

All items on the UK NSC Reporting Checklist for Evidence Summaries have been addressed in this report. A summary of the checklist, along with the page or pages where each item can be found in this report, is presented Table 16.

Table 16. UK NSC reporting checklist for evidence summaries

	Section	Item	Page no.
1.	TITLE AND SUMMARIES		
1.1	Title sheet	Identify the review as a UK NSC evidence summary.	Title page
1.2	Plain English summary	Plain English description of the executive summary.	5
1.3	Executive summary	Structured overview of the whole report. To include: the purpose/aim of the review; background; previous recommendations; findings and gaps in the evidence; recommendations on the screening that can or cannot be made on the basis of the review.	7
2.	INTRODUCTION AND APPROACH		
2.1	Background and objectives	Background – Current policy context and rationale for the current review – for example, reference to details of previous reviews, basis for current recommendation, recommendations made, gaps identified, drivers for new reviews	10
		Objectives – What are the questions the current evidence summary intends to answer? – statement of the key questions for the current evidence summary, criteria they address, and number of studies included per question, description of the overall results of the literature search.	12
		Method – briefly outline the rapid review methods used.	15
2.2	Eligibility for inclusion in the review	State all criteria for inclusion and exclusion of studies to the review clearly (PICO, dates, language, study type, publication type, publication status etc.) To be decided *a priori*.	17

2.3	Appraisal for quality/risk of bias tool	Details of tool/checklist used to assess quality, e.g. QUADAS 2, CASP, SIGN, AMSTAR.	19
3.	SEARCH STRATEGY AND STUDY SELECTION (FOR EACH KEY QUESTION)		
3.1	Databases/ sources searched	Give details of all databases searched (including platform/interface and coverage dates) and date of final search.	19
3.2	Search strategy and results	Present the full search strategy for at least one database (usually a version of Medline), including limits and search filters if used. Provide details of the total number of (results from each database searched), number of duplicates removed, and the final number of unique records to consider for inclusion.	Appendix 1
3.3	Study selection	State the process for selecting studies – inclusion and exclusion criteria, number of studies screened by title/abstract and full text, number of reviewers, any cross checking carried out.	15
4.	STUDY LEVEL REPORTING OF RESULTS (FOR EACH KEY QUESTION)		
4.1	Study level reporting, results and risk of bias assessment	For each study, produce a table that includes the full citation and a summary of the data relevant to the question (for example, study size, PICO, follow-up period, outcomes reported, statistical analyses etc.). Provide a simple summary of key measures, effect estimates and confidence intervals for each study where available. For each study, present the results of any assessment of quality/risk of bias.	Appendix 3
5.	QUESTION LEVEL SYNTHESIS		
5.1	Description of the evidence	For each question, give numbers of studies screened, assessed for eligibility, and included in the review, with summary reasons for exclusion.	21,28,35,42
5.2	Combining and presenting the findings	Provide a balanced discussion of the body of evidence which avoids over reliance on one study or set of studies. Consideration of four components should inform the reviewer's judgement on whether the criterion is 'met', 'not met' or 'uncertain': quantity; quality; applicability and consistency.	22,29,35,42
5.3	Summary of findings	Provide a description of the evidence reviewed and included for each question, with reference to their eligibility for inclusion.	26,33,41,45

		Summarise the main findings including the quality/risk of bias issues for each question. Have the criteria addressed been 'met', 'not met' or 'uncertain'?	
6.	REVIEW SUMMARY		
6.1	Conclusions and implications for policy	Do findings indicate whether screening should be recommended? Is further work warranted? Are there gaps in the evidence highlighted by the review?	46
6.2	Limitations	Discuss limitations of the available evidence and of the review methodology if relevant.	47

References

[1] UK National Screening Committee. Screening for osteoporosis in postmenopausal women. Briefing Note, August 2018
[2] Peto L. Allaby M. Screening for osteoporosis in postmenopausal women: a report for the UK National Screening Committee. January 2013
[3] National Institute of Health and Care Excellence. Osteoporosis: assessing the risk of fragility fracture. Clinical Guideline 146, 2012 (last updated in February 2017)
[4] Adler RA. Osteoporosis in men: a review. Bone Research 2014, 2: 14001
[5] Compston J. Cooper A. Cooper C. et al. (National Osteoporosis Guideline Group). UK Clinical guideline for the prevention and treatment of osteoporosis. Arch. Osteoporosis 2017, 12: 43
[6] National Institute for Health and Care Excellence. Alendronate, etidronate, risedronate, raloxifene and strontium ranelate for the primary prevention of osteoporotic fragility fractures in postmenopausal women. NICE Technology Appraisal Guidance 160 (amended), 2011
[7] National Institute for Health and Care Excellence. Denosumab for the prevention of osteoporotic fractures in postmenopausal women. NICE Technology Appraisal Guidance 204, 2010
[8] Viswanathan M. Reddy S. Berkman N. et al. Screening to prevent osteoporotic fractures: an evidence review for the U.S. Preventative Services Task Force. Evidence Synthesis No. 162. AHRQ Publication No. 15-05226-EF-1. Rockville, MD: Agency for Healthcare Research and Quality; 2018
[9] Viswanathan M. Reddy S. Berkman N. et al. Screening to prevent Osteoporotic Fractures: Updated evidence report and systematic review for the US Preventative Services Task Force. JAMA 2018, 319(24): 2532-2551
[10] Nelson HD. Haney EM. Dana T. Bougatsos C. Chou R. Screening for osteoporosis: an update for the US Preventative Services Task Force. Annals of Internal Medicine 2010, 153(2): 99-111
[11] Rubin KH. Rothmann MJ. Holmberg T. et al. Effectiveness of a two-step population-based osteoporosis screening program using FRAX: the randomized Risk-stratified Osteoporosis Evaluation (ROSE) study. Osteoporosis International 2018, 29: 567-578
[12] Shepstone L. Lenaghan E. Cooper C. et al. Screening in the community to reduce fractures in older women (SCOOP): a randomised controlled trial. Lancet 2018, 391: 714-747
[13] Turner DA. Khioe RFS. Shepstone L. et al. The Cost-Effectiveness of Screening in the Community to Reduce Osteoporotic Fractures in Older Women in the UK: Economic Evaluation of the SCOOP Study. Journal of Bone & Mineral Research 2018, 33(5): 845-851