Romana Carisa Carvalho Vieira

Screening for low bone mineral density in men

Porto | 2013

Dissertação de candidatura ao grau de Mestre em Epidemiologia, apresentada à Faculdade de Medicina da Universidade do Porto, realizada sob a orientação científica da Doutora Raquel Lucas Calado Ferreira, co-orientada pela Professora Doutora Carla Maria de Moura Lopes (Departamento de Epidemiologia Clínica, Medicina Preventiva e Saúde Pública da Faculdade de Medicina da Universidade do Porto e Instituto de Saúde Pública da Universidade do Porto).

Esta investigação foi realizada no âmbito do estudo EPIPorto (coordenado pelo Professor Doutor Henrique Barros), especificamente do projeto *Peripheral body fat, lifestyles and adipokines*, financiado pela Fundação para a Ciência e a Tecnologia, Portugal [PTDC/SAU-ESA/108315/2008].

U.PORTO
FMUP FACULDADE DE MEDICINA
UNIVERSIDADE DO PORTO

Ao abrigo do Art.º 8º do Decreto-Lei nº 288/70 esta dissertação teve como base dois manuscritos, nos quais colaborei ativamente na definição das hipóteses, recolha, análise e interpretação dos dados. Fui responsável pela redação da versão inicial dos dois manuscritos:

I. Clinical screening tools to identify men with low bone mass: a systematic review.

II. Upper arm circumference measurement improves screening for low bone mineral density in men.

Table of Contents

Agradecimentos

A Medicina não será mais a mesma para mim, não só porque os dois últimos anos me trouxeram novos conhecimentos mas, essencialmente, porque a forma como agora vejo a informação que me é dada mudou.

Mas o lucro dos dias passados vai para além da própria ciência e refina-se nas pessoas que me deu a conhecer. É a esses que gostaria de aqui deixar a minha gratidão.

Em primeiro lugar quero agradecer às minhas orientadoras. À Doutora Raquel Lucas, a confiança, oportunidades e inspiração que me concedeu. À Professora Doutora Carla Lopes pelos conhecimentos transmitidos desde há muitos anos e que serviram de mote a esta etapa.

Aproveito ainda para agradecer ao Professor Milton Severo e ao Fábio Araújo pelos seus contributos no desenvolvimento desta tese.

Não posso deixar de agradecer a todos os restantes membros do Instituto de Saúde Pública da Universidade do Porto pela disponibilidade que mostraram quando solicitados.

Às minhas colegas de Mestrado, amigas nas horas de euforia mas também nas de incerteza.

Quero ainda deixar o meu agradecimento aos meus colegas, internos de Reumatologia, pelo apoio que sempre me deram, por vezes com prejuízo pessoal.

E aos meus amigos de sempre e para sempre por estarem lá quando preciso.

Ao Bé, por perceber e aceitar que amar nem sempre é estar presente. Sem o seu apoio tudo seria bem mais difícil.

Ao meu irmão, pelo exemplo que se tornou na minha vida.

Aos meus pais, a quem admiro acima de todos, pelos sacrifícios que fizeram por nós.

E à minha mãe, que sempre acreditou em mim. É ela quem hoje procuro na espuma das ondas, onde sempre a encontro, com um travo de sal.

RESUMO

Introdução: As fraturas de fragilidade óssea são uma causa importante de morbimortalidade em todo o mundo. Embora mais frequentes nas mulheres, nos homens as fraturas da anca associam-se a pior prognóstico. No entanto, a fragilidade óssea masculina continua a ser um tema bastante negligenciado. Embora a determinação da densidade mineral óssea (DMO) seja um passo chave na avaliação da fragilidade óssea, o rastreio universal não é eficiente pelo que, alternativamente, se deva testar apenas indivíduos com maior probabilidade de ter baixa DMO. Diversas ferramentas de decisão clínica foram desenvolvidas com o objetivo de identificar esses indivíduos mas a sua validação em diferentes populações levou a grande heterogeneidade nos limites propostos para positividade, tornando a sua aplicabilidade confusa e alertando para necessidade da calibração das mesmas. Por fim, a maioria destas ferramentas é baseada em cálculos complexos pelo que, desde que assegurada validade semelhante, quanto mais simples for a ferramenta, mais provável será a sua adoção na prática clínica.

Objetivos: Estudar a capacidade das ferramentas de decisão clínica identificarem homens com baixa DMO que beneficiem da realização de densitometria óssea, através de dois objetivos específicos: 1) descrever e comparar a validade das ferramentas de decisão clínica publicadas para identificar homens com baixa DMO através da elaboração de uma revisão sistemática; 2) calibrar as ferramentas OST e MORES para a nossa população e avaliar a sua capacidade discriminatória em relação a uma nova ferramenta clínica (EPIPOST), desenvolvida e validada no presente estudo, para identificar homens com baixa DMO.

Objetivo 1: Na elaboração da revisão sistemática foram identificados, em base de dados eletrónica (Medline) e na lista de referências dos artigos incluídos, estudos que descrevessem a validade de instrumentos de predição de baixa DMO em homens. Após uma primeira seleção por leitura dos títulos e resumos de todas as citações, foi feita uma segunda seleção por leitura completa dos artigos selecionados. Foram colhidos dados de forma estruturada a partir dos artigos selecionados relativamente a características dos participantes, especificidades da densitometria óssea e propriedades da ferramenta validada em termos de fatores de risco incluídos e medidas de capacidade discriminativa calculadas. A qualidade metodológica foi avaliada através de uma versão modificada da lista QUADAS.

Foram identificadas 1484 citações: 1447 foram excluídas após leitura do título e resumo uma vez que não cumpriam os critérios de inclusão. Os restantes 37 artigos foram lidos integralmente. No final, foram incluídos 22 artigos: 2 estudos avaliaram a capacidade de 5 *guidelines* para realização de densitometria óssea propostas por diferentes entidades; 5 estudos desenvolveram e validaram 5 novas ferramentas de rastreio mas destas apenas 2 foram posteriormente validadas noutras populações; 12 estudos avaliaram a capacidade do OST, 4 a do MORES e 3 a do OSTA. Verificou-se grande heterogeneidade entre os estudos [em termos de características da amostra (população de base, idade, raça); medição da massa óssea (equipamento de densitometria e avaliação da qualidade), diagnóstico de baixa DMO (população de referência para o cálculo do T-score, local anatómico selecionado)] e a qualidade global foi moderada tal como evidenciado por uma média de 10.8 itens, com variação entre 8 e 15, em 19 possíveis na lista QUADAS modificada. Embora nenhuma das ferramentas se tenha mostrado consistentemente melhor que a outra nem houvesse consenso relativamente ao melhor valor limite para positividade, todas mostram capacidade preditiva razoável pelo que se deve optar pela de mais simples execução.

Objetivo 2: Como parte do estudo de base populacional em adultos EPIPorto foram avaliados 147 homens entre os 40 e os 65 anos. Foram registados idade, altura, peso, índice de massa corporal e diversas circunferências corporais. Para avaliação da DMO foi realizada densitometria de corpo inteiro. Para a calibração das ferramentas OST e MORES, foram estimados novos parâmetros de regressão baseados nas caraterísticas da nossa população. No desenvolvimento da nova ferramenta de rastreio EPIPOST, as diferentes medidas antropométricas para prever baixa DMO foram testadas através de diferentes modelos de regressão logística. A validação do EPIPOST foi realizada pelo método *leave-one-out cross-validation*. O ajuste global e capacidade discriminatória foram testados por comparação direta dos valores previstos e estimados de baixa DMO por quartil de pontuação de cada ferramenta, pelo teste de *"goodness-of-fit"* Hosmer-Lemeshow e pela área sob a *receiver operating characteristic curve* (AUC). Finalmente, foram calculados *likelihood ratios* (LR) para selecionar os limites de positividade mais adequados para cada ferramenta.

A calibração manteve a capacidade discriminatória do OST e do MORES (AUC 0.73 e 0.75, respetivamente) e melhorou o ajuste dos modelos à nossa população. O EPIPOST, que incluiu apenas a circunferência do braço relaxado, mostrou ligeiramente melhor capacidade discriminatória (AUC 0.76) que as outras ferramentas. A análise de LR mostrou que o EPIPOST apresenta maior capacidade discriminativa ao longo dos

diferentes níveis de risco (LR a variar entre 0.1 e 18.4, em comparação com 0.0 e 2.4 com o OST e 0.2 e 2.8 com o MORES).
Para prever baixa DMO, um OST≤2 teve sensibilidade de 100% e especificidade de 8.2%; um MORES>-2 teve sensibilidade de 93.9% e especificidade de 30.6%; um EPIPOST>-2 teve sensibilidade de 98.0% e especificidade de 18.6%. A validação do EPIPOST revelou que a concordância entre os valores observados e previstos foi aceitável (73.3% na amostra de desenvolvimento e 71.2% nas amostras de validação).

Conclusões: A nossa revisão sistemática identificou 22 estudos que validavam 9 ferramentas. No entanto, destas, apenas 3 (OST, OSTA e MORES) foram validadas mais que uma vez em populações masculinas. Foi constatada elevada heterogeneidade metodológica entre os estudos identificados e não houve concordância quanto ao limite para positividade mais adequado para cada ferramenta. Ainda assim, todas as ferramentas mostraram razoável capacidade preditiva e nenhuma foi superior às demais, pelo que o OST e o MORES, dada a sua simplicidade, capacidade preditiva e replicação, parecem ser mais adequados para uso na prática clínica. A calibração das ferramentas OST e MORES melhorou o ajuste de ambos os modelos à nossa população embora a capacidade discriminatória para identificar homens com baixa DMO se tenha mantido. A ferramenta desenvolvida por nós especificamente para homens, o EPIPOST, mostrou ter melhor capacidade discriminativa que o OST e que o MORES, bem como ser mais fácil de executar. A análise de *likelihood ratios* mostrou que indivíduos que pontuassem OST≤2, MORES> -2 ou EPIPOST>-2 tinham maior probabilidade de ter baixa DMO e, portanto, deviam realizar densitometria óssea.

ABSTRACT

Introduction: Bone fragility and associated fractures are an important cause of morbidity and mortality worldwide. Even though lifetime risk of hip fracture is higher in women, fragility fractures are associated with worst prognosis in men. Nevertheless, male osteoporosis remains a neglected condition. Although bone mineral density (BMD) determination is a key step for bone fragility assessment, mass screening is not cost-effective. An alternative involves testing only individuals with higher probability of low BMD. Several clinical decision rules have been developed with this objective but their validation across different populations has led to heterogeneity in the proposed cut-offs, making its applicability unclear and calling for calibration. Finally, most of these tools are based on complex calculations. However, when similar validity can be ensured, the simpler the clinical screening tool the more likely it is to be applied in clinical practice.

Aims: To study the capacity of clinical decision rules to identify men with low BMD who should undergo dual-energy x-ray absorptiometry (DXA) testing through the following specific objectives: 1) to describe and compare the validity of published clinical screening tools designed to identify men with low bone mineral density trough the elaboration of a systematic review; 2) to calibrate OST and MORES tools to our population and to assess their discriminatory ability against a newly developed and validated simplified clinical screening test (EPIPOST) to identify men with low BMD.

Objective 1: For the systematic review elaboration, we identified studies describing the validation of instruments for low bone mass prediction in men in electronic database (Medline) and reference lists of included articles. After screening all titles and abstracts, the selected articles were rescreened by full text reading. Data were extracted on participants' characteristics, DXA features and tools' properties in terms of risk factors included and discriminatory performance of all selected studies. Methodological quality was assessed using a modified QUADAS checklist.

A total of 1484 citations were identified with our search: 1447 were excluded after screening the title and abstract as they did not met our inclusion criteria. The remaining 37 articles were submitted to full-text reading. At the end, 22 articles were included in our systematic review: 2 studies assessed the performance of 5 guidelines for DXA testing proposed from different entities; 5 studies developed and validated 5 new screening tools but of these only 2 were further validated in other populations; 12 studies

evaluated the performance of OST, 4 assessed the performance of MORES and 3 estimated the performance of OSTA. There was high heterogeneity across studies [in terms of sample characteristics (source population, age, race); bone mass assessment (DXA equipment and quality assessment), low bone mass diagnosis (reference population for T-score calculation, anatomical site selected)] and the global quality was moderate as showed by a mean of 10.8 items, with a range of 8-15, in 19 possible modified QUADAS items. Despite substantial variability regarding the most accurate cut-off, all tools showed moderate predictive capacity and any performed consistently better than other, denoting that the simplest to execute should be preferred.

Objective 2: As part of EPIPorto population-based study among adults, 147 men aged between 40 and 65 years were assessed. Age, height, weight, body mass index and several body circumferences were recorded. DXA whole-body scans were performed for BMD assessment. For OST and MORES calibration, new regression parameters were estimated based on our population features. For EPIPOST development, the different anthropometric variables were tested using logistic regression models to predict low BMD. EPIPOST validation was done by the leave-one-out cross-validation method. The overall fit and discriminatory capacity were assessed by direct comparison of the observed and expected prevalences of low BMD by quartiles of each score, Hosmer-Lemeshow "goodness-of-fit" test and area under the receiver operating characteristic curve (AUC). Finally, likelihood ratios (LR) were calculated to select the ideal cut-off for each model.

Calibration maintained the discriminatory capacity of OST and MORES (AUC 0.73 and 0.75 respectively) but improved the fitting. EPIPOST included only relaxed upper arm circumference and showed slightly better discriminatory capacity (AUC 0.76) than the other tools. The LR analysis showed that EPIPOST had higher discriminative ability across different risk levels (LR range of 0.1 to 18.4, compared to 0.0 to 2.4 with OST and 0.2 to 2.8 with MORES). For predicting low BMD, OST≤2 had a sensitivity of 100% and a specificity of 8.2%; MORES>-2 had a sensitivity of 93.9% and a specificity of 30.6%; EPIPOST>-2 had a sensitivity of 98.0% and a specificity of 18.6%. EPIPOST leave-one-out cross-validation showed that the agreement between the observed and predicted values was acceptable (73.3% in the training sample and 71.2% across validation samples).

Conclusions: Our systematic review identified 22 studies validating 9 tools aimed to identify men with low BMD more likely to benefit from DXA testing. However, of these, only 3 (OST, OSTA and MORES) were validated in more than one sample. Despite high

methodological heterogeneity across studies and uncertainty regarding the most accurate cut-off, all tools showed reasonable predictive capacity and any performed better than the others. OST and MORES, for their simplicity, accuracy and replication, seem to be more adequate for routine clinical practice use. Calibration of OST and MORES improved the fitting of both models to our population data while maintaining their discriminatory ability to identify men with low BMD. The newly developed tool specifically for men, EPIPOST, proved to have even better discriminatory ability than OST and MORES despite being simpler to execute. The likelihood ratios analysis revealed that individuals scoring OST≤2, MORES>-2 and EPIPOST>-2 should undergo DXA testing.

INTRODUCTION

Condition

Osteoporosis is a systemic skeletal disease characterized by reduced bone mass and microarchitectural deterioration of bone tissue that increases bone fragility and risk of fractures [1]. Low-energy fractures, also known as fragility fractures, are the main complication of the condition. A fragility fracture results from forces that would not normally cause a fracture, such as a hip or wrist trauma from falling from standing height or low-energy vertebral compression. Although specific fracture sites have been considered more characteristic of osteoporosis, fractures occurring at nearly every anatomical site have been associated with bone fragility [2].

Prevalence and Burden of Osteoporosis

Osteoporosis and osteoporotic fractures are rare before the age of fifty years. However, their prevalence rises with age thereafter and it was estimated that in Europe approximately 6% of all men and 21% of all women aged 50–84 years have osteoporosis [3]. It has been also estimated that, between 2010 and 2050, the prevalence of osteoporosis will increase by 40% in women and 81% in men and fragility fractures will raise by 218% in women and 335% in men [4].

More importantly, the incidence of non-vertebral fractures has been estimated of 9.6 (95% CI, 8.3 - 11.0) per 1000 person-years in men and 25.0 (95% CI, 23.3 - 26.9) per 1000 person-years in women [5]. Information about vertebral fractures is more difficult to deliver as they are often subclinical and usually managed in ambulatory and, therefore, remain unidentifiable through hospital databases.

Amongst patients who have experienced fractures, it is well documented that fewer than 50% are ever assessed or treated for osteoporosis [6]. This has major significance as fragility fractures are associated with an increase in mortality [7]. It was estimated that approximately 34,000 deaths annually are caused by fractures in Europe [3].

However, in terms of disease burden, it is more important to consider the excess mortality due to fragility fractures. Excess mortality is substantial after hip fracture, depending on the time since the event (higher in the immediate fracture period untill the end of the first year) but also on age and sex of the individuals [8].

Specific aspects of osteoporosis in men

Even though the lifetime risk of hip fracture was estimated as 8.9% (95% CI, 2.3 - 15.4%) in women and 6.7% (95% CI, 1.2 - 12.2%) in men [9] it is known that fragility fractures denote worst prognosis in the male sex: the cumulative mortality at 12 months among individuals with hip fracture patients compared to the general population was 37.1% in men and 26.4% in women [10]. The reasons relate to the higher general population mortality in men but also with the higher relative risk of death after fracture in men compared with women [3].

Nevertheless, male osteoporosis remains a neglected condition: Papaioannou et al. [11] reported a diagnostic and therapeutic care gap in a sample of men with fragility fractures whit only 2.4% of these reporting an osteoporosis diagnosis. For many years, osteoporosis has been considered to be a condition associated with postmenopausal women. This misconception has led to underreferral of men for dual-energy X-ray absorptiometry (DXA) and, consequently, underdiagnosis of osteoporosis even in the presence of fragility fractures.

In terms of clinical research, there is also a paucity of reported clinical studies exploring the low bone mineral density (BMD) in men, for example, in terms of reports studying the efficacy of osteoporosis therapy, when compared with large studies conducted in women [12].

Diagnosis

Osteoporosis is diagnosed in individuals on the basis of presence of a fragility fracture or by bone mass measurement criteria [2].

The definition of an osteoporotic fracture is not straightforward and there are diverging opinions on what to consider an osteoporotic fracture. One approach is to consider those resulting from low energy trauma, defined as a fall from a standing height or less, or trauma that in a healthy individual would not give rise to fracture [13]. Data from a systematic review state that the fractures most likely related to osteoporosis were those at the femoral neck and lumbar and thoracic vertebrae. The fractures rated least likely because of osteoporosis were open proximal humerus fractures, skull, and facial bones [14].

The goal in osteoporosis management is to avoid fracture. Therefore clinicians must seek upstream for asymptomatic patients and assess characteristics that pertain to bone physical properties. The description of osteoporosis captures the notion that low bone mass is an important component of the risk of fracture, but other abnormalities occur in the skeleton that contribute to skeletal fragility. Indeed, bone strength is determined by bone mineral density, cortical thickness and porosity, trabecular bone morphology, and intrinsic properties of bony tissue [15]. Ideally, the assessment of the skeleton should capture all these aspects. However, bone mass and area (and their composite measure bone mineral density) remain the parameters routinely measured in clinical practice despite capturing only a part of overall bone strength. Nevertheless, it has been estimated that BMD measured by DXA accounts for 60-70% of bone strength variability in the population [15].

Bone mineral density criteria for the definition of osteoporosis were proposed by the World Health Organization (WHO) from epidemiologic data that describe the normal distribution of BMD in a young healthy reference population [16]. BMD is the amount of bone mass per unit volume (volumetric density, g/cm^3), or per unit area (areal density, g/cm^2), and both can be measured in vivo. A large variety of techniques is available but the most widely used techniques by far are based on X-ray absorptiometry in bone, particularly DXA that measures areal BMD. The distribution of bone mineral content or density in young healthy adults (representing “peak bone mass”) is approximately normal, irrespective of the measurement technique used. Because of this normal distribution, bone density values in individuals may be expressed in relation to a reference population in standard-deviation (SD) units. When SDs are calculated in relation to the mean of a young healthy population, this is referred to as the T-score [3]. It has been estimated that for each SD decrease in BMD, measured by DXA, the risk of hip fracture increases 2.6-fold [17]. This risk gradient led, in 1994, to the publishing of

the WHO osteoporosis diagnostic criteria based on the T-score measured at the hip, spine or forearm, with T-score≥-1.0 being considered normal, -2.5<T-score<-1.0 considered osteopenia and T-score≤-2.5 considered osteoporosis [16]. More recently, the femoral neck has been adopted as the standard measurement site and the reference population for both men and women T-score calculation being the mean and SD values in young women from the NHANES III study [18].

Screening for osteoporosis

Even though BMD measurement by DXA is considered the gold standard assessment for determining who has low BMD, screening the entire population of men with DXA scans has very low efficiency due to low pre-test probability, suboptimal DXA availability and elevated costs.

A worthwhile alternative involves identifying men with higher probability of low BMD who will subsequently be suitable candidates for DXA screening, as determined through screening tools composed by risk factors for low BMD. These screening tools increase DXA screening efficiency and cost-effectiveness by reducing the number of men referred who are otherwise healthy.

Ideally, a screening test should be 100% sensitive and 100% specific. However, in practice, this does not occur as sensitivity and specificity are usually inversely related. A test with good sensitivity is favoured when a false-negative result is more prejudicial for the patient than a false-positive one (curable disease, early diagnosis associated with better prognosis) or when the disease is uncommon. A test with a good specificity is favoured when a false-positive result is more prejudicial for the patient than a false-negative one (aggressive treatment, incurable disease, condition with heavy labelling effect, etc.) [19].

In case of a false-positive result from an osteoporosis screening tool, there is no risk of harm to the patient from unnecessary treatment or invasive diagnostic testing because treatment for low BMD would only be initiated upon confirmation by DXA, a safe and non-invasive diagnostic procedure. In this way, more importance has to be given to sensitivity rather than specificity when developing a clinical decision rule in this field. In

terms of likelihood ratios, this means that screening tests should be associated to low negative likelihood ratios.

Screening in men

Several screening strategies have been proposed to identify men who benefit the most from DXA testing. Worldwide, different guidelines have been written for the diagnostic evaluation of male osteoporosis but they are not as well validated as in postmenopausal osteoporosis.

The 2010 NOF (National Osteoporosis Foundation) Clinician's Guide recommends DXA testing in men with specific conditions (e.g., rheumatoid arthritis) or on specific medications (e.g., glucocorticoids in a daily dose ≥5 mg prednisone or equivalent for ≥3 months) that may predispose to bone loss, in men age 70 and older regardless of clinical risk factors as well as in men aged 50-70 when they have a prior fragility fracture or because of clinical risk factors [20]. The Portuguese Society of Rheumatology also recommends that all men above 70 years should undergo DXA testing and that the remaining should only be tested depending on the presence of other risk factors for low BMD [21]. However, the accuracy of the criteria proposed in these recommendations for selecting patients to DXA has not been estimated and these guidelines also lack cost-effectiveness assessment. In fact, despite widespread screening of men older than 70 years has been claimed, Schousboe et al. [22] reported that universal screening would probably be cost-effective only in men aged 80 years or older.

These facts highlight the necessity of a tailored approach in screening for low BMD in men. Instead of empirical guidelines based on risk factors associated with low BMD, it has been proposed that clinicians should follow clinical decision rules that have been validated, an option which may require several studies to fully test the accuracy of the rule in different clinical settings [23].

In the last twenty years, several clinical decision rules have been developed with the objective of identifying individuals with low BMD who should undergo DXA testing. They comprise various components of the medical history, physical examination, laboratory results and other complementary exams (e.g., ultrasound). As a whole, clinical decision

rules inform clinical judgment and have the potential to change clinical behaviour and reduce unnecessary costs while maintaining quality of care. [23].

Clinical screening tools

Specifically, decision rules based only on variables from the medical history and physical examination (clinical screening tools) have many advantages as they provide more readiness in the information availability and avoid costs, time consumption and adverse events (for example in the case of ionizing radiation) related to additional complementary exams.

The implementation of clinical screening tools is being attempted in many medical areas, mainly because they represent a way to improve the quality of care while reducing health costs.

Although there are few studies of the cost-effectiveness of clinical screening tools in osteoporosis, Shepherd et al. [24] estimated that the clinical utility of Male Osteoporosis Risk Estimation Score (MORES), as measured by the number needed to screen (NNS) to prevent 1 additional hip fracture, compared favourably with the United States Preventive Services Task Force (USPSTF) findings for women: USPSTF recommended universal DXA testing for women aged 65 years with a NNS of 731 to prevent 1 additional hip fracture; MORES, with similar assumptions of subsequent treatment, adherence, and fracture reduction in those diagnosed with osteoporosis, resulted in a NNS to prevent 1 additional hip fracture of 544 in men aged 60 to 64 years.

Since the mid-1990s several articles appeared in the literature regarding pre-screening instruments to select patients for DXA testing. Each of these instruments focused on different characteristics, sometimes with overlap, to reach different decision points with different cut offs. Specifically in men, a meta-analysis from 2008 [25] found 5 articles concerning clinical screening tools to identify men with low BMD. Since then, new tools were developed and a number of papers to assess the accuracy of previously developed clinical screening tools in different populations has been published.

The validation of these screening tools across different populations has led to heterogeneity in values selected as screening thresholds at which DXA is recommended [25], making its applicability in clinical practice confusing. This fact calls for the need to calibrate these tools for specific populations rather than only constantly change the decision thresholds.

Finally, most of the developed screening tools are based on complex calculations using self-reported variables such as age and weight. However, self-reported weight is known as a subject of important information bias when compared to measured weight, particularly in men [26]. Therefore, when similar validity can be ensured, the simpler the clinical decision rule the more likely it is to be applied in real clinical settings.

AIMS

The aim of this thesis was to study the capacity of clinical decision rules to identify men with low BMD who should undergo DXA testing through the following specific objectives:

1. To describe and compare the validity of published clinical screening tools designed to identify men with low bone mineral density trough the elaboration of a systematic review.
2. To calibrate OST and MORES tools to our population and assess their discriminatory ability against a newly developed and validated simplified clinical screening test (EPIPOST) to identify men with low BMD.

CHAPTER I

Clinical screening tools to identify men with low bone mass: a systematic review

Romana Vieira, Fábio Araújo, Carla Lopes, Raquel Lucas

Abstract

Objective: To identify and assess the accuracy of published clinical screening tools designed to identify men with low bone mineral density (BMD). **Methods:** We identified studies describing the validation of instruments for low bone mass prediction in men through an electronic database search (Medline), including reference lists of eligible articles. We screened for inclusion first by reading the title and abstract of all retrieved references and then by full text reading of the articles selected in the first phase. Data were extracted on participants' characteristics, dual-energy x-ray absorptiometry (DXA) features and tools properties in terms of risk factors included and discriminatory performance. Methodological quality was assessed using a modified QUADAS checklist. **Results:** A total of 1484 citations were identified with our search strategy: 1447 were excluded after screening the title and abstract as they did not met our inclusion criteria. The remaining 37 articles were submitted to full-text reading. At the end, 22 articles, were included in our systematic review: 2 studies assessed the performance of 5 guidelines for DXA testing proposed by different entities; 5 studies developed and validated 5 new screening tools but of these only 2 were further validated in other populations; 12 studies evaluated the performance of OST, 4 assessed the performance of MORES and 3 estimated the performance of OSTA. There was high heterogeneity across studies (regarding source populations, age distribution, ethnic background, low BMD diagnostic criteria, DXA equipment) and the global quality of the reports was moderate as shown by a mean of 10.8 items, with a range of 8-15, in 19 possible modified QUADAS items. No tool performed consistently better than the others and none of the cut-offs tested was evidently more accurate than the alternatives for each score. **Conclusion:** Nine instruments were validated in 22 studies. However, only 3 (OST, OSTA and MORES) had been validated in more than in one sample. Overall, even though no clear cut-offs for any of the scores emerged from our analysis, all tools had acceptable predictive capacity and performed similarly in terms of overall accuracy. OST and MORES, for their simplicity, accuracy and replication, seem to be more adequate for routine clinical practice use.

Keywords: Men, Bone Density, Screening, Clinical Prediction Rule, Systematic Review.

Introduction

Osteoporosis is a disease defined by decreased bone mass and alteration of bone micro architecture which results in increased bone fragility [1] with a growing incidence worldwide. Although it mostly affects postmenopausal women, it is known that fracture, its major consequence, has worse prognosis in men regarding not only morbidity but also mortality [2].

Since bone mass assessment using dual-energy X-ray absorptiometry (DXA) is the gold standard method to diagnose osteoporosis [3], several clinical decision rules (CDR) have been developed to help clinicians deciding who should undergo DXA examination. A CDR quantifies the independent contribution of each component of the clinical history, physical examination and basic laboratory results towards a specific condition in an individual patient [4] with the objective of increasing the clinician´s diagnosis accuracy through a case-finding strategy. In the bone health field, since the mid-1990s, clinical screening tools have been developed to select patients for bone densitometry [5]. However, at present, there is no universally accepted policy for population screening in Europe to identify patients with osteoporosis or those at high risk of fracture.

Most osteoporosis screening tools have been developed in women and some of those were subsequently validated in men. More recently, attending for gender heterogeneity in terms of candidate bone mineral density (BMD) predictors, some groups developed male-specific osteoporosis screening tools. However, none of them has found broad acceptance in medical practice so far and there is no wide consensus on their recommended cut-off values. The election of a specific tool in detriment of others would be enhanced if one proved to be more accurate in terms of predictive ability and/or easier to apply, such as composed solely of clinical risk factors. However, to the best of our knowledge, the only synthesis of published evidence on CDR in men dates from 2008, even though a number of new validation studies have been conducted and published since then [6].

The objective of the present systematic review was to identify and assess the accuracy of published clinical screening tools designed to identify men with low bone mineral density.

Methods

We followed the current analytical methods and standards established by the Preferred Reporting Items for Systematic Reviews and Meta‐Analyses (PRISMA) group for systematic reviews and meta-analysis [7].

Eligibility criteria

We included articles written in English, French, Spanish, Italian and Portuguese.

- Population: adult male.
- Intervention: validation of prognostic instruments to identify individuals with an increased risk of low bone mineral density - development of risk assessment tools was included if they were derived from an initial population and then validated internally or externally). Due to the practical value of the tool, we have considered scores including only clinical risk factors (not derived from complementary diagnostic exams).
- Included studies had to report characteristics of the screening tool performance [(sensitivity, specificity or area under the receiver operating characteristic curve (AUC)].
- Outcome: bone mineral density measured by DXA at femoral neck, total hip or lumbar spine.

We excluded case reports, non-systematic reviews, letters to the editor, editorials and commentaries. Screening tools developed in populations defined based on specific diagnosis were also excluded even though we accepted studies on screening tools developed in population based samples and subsequently validated in specific populations.

Search

The studies were identified by searching Medline (PubMed). The final electronic search was undertaken on August 27, 2013. Additionally, we hand-searched the reference list of all relevant studies (including two previous reviews).

To generate search terms we conducted a preliminary search using known terms and synonyms suggested by clinicians and experts in the field to reach all possible words

related to our study objective. Additionally we added terms and keywords found in relevant studies to augment our search sensitivity.

We used the following search strategy in MEDLINE (PubMed): *(osteoporosis OR osteopenia OR bone mass) AND (instrument OR questionnaire OR models OR algorithm OR scores OR tool) AND (((((((validity OR validation OR "sensitivity and specificity"[All Fields]) OR "specificity"[All Fields]) OR "screening"[All Fields]) OR "false positive"[All Fields]) OR "false negative"[All Fields]) OR "accuracy"[All Fields])) OR ((((("predictive value"[All Fields] OR "predictive value of tests"[All Fields]) OR "predictive values"[All Fields])) OR ("reference value"[All Fields] OR "reference values"[All Fields]) OR (((((("roc"[All Fields] OR "area under curve" [All Fields]) OR "roc auc"[All Fields]) OR "roc curve"[All Fields]) OR "roc curves"[All Fields]) OR "likelihood ratio"[All Fields])))) NOT (animals NOT humans) AND (male or men)*.

Study selection

The first eligibility assessment was conducted by one researcher and consisted in screening the title and abstract of all identified records. Subsequently, two researchers including the one who initially screened the references performed a full text read of potentially relevant studies. At this point disagreements between reviewers were resolved by consensus.

Data collection process

For data collection we developed and applied an extraction sheet to the final list of selected articles for inclusion in the systematic review. This extraction sheet was based on others mentioned in earlier systematic reviews on osteoporosis screening tools in women [8]. On each study we extracted data on: participant's characteristics (number, age, geographical location and clinical context); screening tool features (variables included, development strategy, statistical methods); outcome assessment (DXA equipment and method and site of BMD measurement). In case of studies developing new screening that tools present data from development and validation samples, we extracted only data from the validation sample.

For every study, sensitivity, specificity and AUC were extracted as the principal measures of the performance of screening tools. Positive and negative likelihood ratios (LR) were calculated for each proposed cut-off.

Risk of bias in individual studies

For assessing risk of bias we used a modified version of the Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS) [9] checklist as proposed in previous studies [10].

Results

Study selection

A total of 1484 citations were identified with our electronic search. The hand search of reference lists did not retrieve any new paper. Of the initial 1484 studies, 1447 were excluded after screening the title and abstract as they did not met our inclusion criteria. The remaining 37 articles were submitted to full-text reading. At the end, 22 articles were included in our systematic review. The study selection process, as well as reasons for exclusion, are summarized in Figure 1.

Studies characteristics

We report performance information of seven tools tested in 22 studies. Each study may have tested more than one osteoporosis screening tool.

As more than one study presented performance data for more than one threshold for low BMD, we present the data from the 22 selected articles in 3 independent tables according to low BMD cut-off: Table 1 presents characteristics of studies on the performance of clinical screening tools to identify men with T-scores≤-2.5; Table 2 refers to T-scores≤-2.0 and Table 3 refers to T-scores≤-1.0.

The sample size ranged from 64 to 6572 individuals and relatively to the source population, only 9 studies were considered population based. The remaining referred to individual from general outpatient clinics (4 studies) or patients from specialty clinics (7 studies). One study did not report source population. Finally, one study assessed the accuracy of the Osteoporosis Screening Tool (OST) in patients with rheumatoid arthritis.

In terms of mean age there was also a wide range of values, from 47.0 to 73.0 years. The most commonly included risk factors in the final tool were age and weight even though other variables as chronic obstructive pulmonary disease, emphysema, gastrectomy, inactivity, personal or family history of fractures and weekly calcium intake were also mentioned.

Most studies reported the DXA equipment utilized but only 17 gave additional information about the specific model. As to the site elected for osteoporosis definition, many studies

reported more than one site. The most prevalent definition was based on either hip or the lumbar spine (14 studies), followed by the hip (10 studies) and finally the lumbar spine (6 studies). One study did not report information about this item. In terms of reference group for the T-score calculation, at the femoral neck, only one study selected the young white women from the National Health and Nutrition Examination Survey (NHANES) population. The remaining selected either young men from NHANES population, the manufacturer's built-in population data or country-specific populations. For the lumbar spine the most frequently used references were derived from young male populations selected by the manufacturer or specific from each country.

Two studies evaluated the performance of previously established clinical practice guidelines to identify men with low BMD [11, 12] and five studies developed and validated 5 new models specifically to identify men with low BMD [13-17]. However, of those, only the Osteoporosis Screening Tool for Asians (OSTA) [18, 19] and the Males Osteoporosis Risk Estimation Score (MORES) [20-22] were further validated in different populations. Logistic regression analysis was the method selected to derive the instrument in four of the five studies that developed new screening tools. However, one study choose to use linear regression analysis [13].

Two osteoporosis screening tools initially developed in women were validated in men: the Khon Kaen Osteoporosis Study (KKOS) [23] and, more extensively, OST [14, 19, 21, 24-32].

Overall, there was high heterogeneity across studies in terms of sample characteristics (source population, age, race); bone mass assessment (DXA equipment and quality assessment), low bone mass diagnosis (reference population for T-score calculation, anatomical site selected). Due to this heterogeneity, we decided not to perform a quantitative synthesis of the results. Nevertheless, enough data were available to elaborate a qualitative comparison between different populations of the performance characteristics of OST [14, 19, 21, 24-32], MORES [15, 20-22] and OSTA [17-19] to predict a T-score≤-2.5 (OST, MORES and OSTA) and a T-score<-2.0 (OST). Tools whose validation was performed in a single population are listed in the tables but not included in the following comparative description, as data on performance reproducibility are warranted to decide on the field implementation of any clinical decision rule.

Performance of screening tools to identify men with T-score<-2.0

When identifying men with T-score<-2.0 at the lumbar spine, femoral neck or total hip, OST estimated AUC ranged from 0.695 to 0.830. Different cut-offs were presented, from OST<-1 to OST<4, with sensitivity and specificity ranging from 62 to 83% and 57 to 89%, respectively. The lowest and highest positive LR were 1.93 and 5.64, respectively. The values for the negative LR ranged from 0.30 to 0.43.

Performance of screening tools to identify men with T-score< -2.5

For OSTA, at the femoral neck, the estimated AUC ranged from 0.848 to 0.850. With OSTA<-1 the estimated sensitivity and specificity ranged from 83 to 91% and 66 to 67%, respectively. At the lumbar spine, femoral neck or total hip, the estimated AUC was 0.780. Different cut-offs were presented, from OSTA<-1 to OSTA<4, with sensitivity and specificity ranging from 38 to 77% and 43 to 82%, respectively. The estimated positive LR ranged from 1.27 to 2.68 and the negative LR from 0.14 to 0.76.

Regarding MORES, at the femoral neck or total hip, the reported AUC estimated ranged from 0.820 to 0.842. A MORES≥6 had a sensitivity ranging from 80 to 95% and a specificity ranging from 61 to 70%. The lowest and highest positive LR were 1.29 and 2.67, respectively. In terms of negative LR, the lowest value was 0.00 and the highest was 0.73.

The performance of OST to predict a T-score<-2.5 at the lumbar spine, femoral neck or total hip, in terms of AUC, ranged from 0.667 to 0.890. The tested cut-offs varied extremely from OST<-3 to OST<8. At the different thresholds, sensitivity ranged from 7 to 95% and specificity from 23 to 99%. When predicting T-score<-2.5 at lumbar spine, AUC ranged from 0.660 to 0.717, sensitivity ranged from 36 to 64% and specificity from 60 to 80%. OST predicted T-score<-2.5 at femoral neck with estimated AUC ranging from 0.740 to 0.990. At the presented cut-offs, sensitivity varied from 6 to 100% and specificity from 51 to 94% with positive LR from 1.00 to 14.00 and negative LR from 0.00 to 1.00. Interestingly, one study found that OST performed much better in African American than Whites (in Whites, the best discriminatory capacity corresponded to OST<4 with a positive and negative LR of 1.73 and 0.29, respectively; in African American, the best discriminatory capacity was found at OST<2, with a positive and negative LR of 5.88 and 0.00).

Finally, a study assessing the performance of OST in men with rheumatoid arthritis found, at the proposed cut-off of OST≤4, a sensitivity of 78% and a specificity of 45% for identifying osteoporosis; for identifying low BMD its sensitivity was 64% and specificity was 54%.

Risk of bias within studies

According to our assessment, the 22 included studies validating tools to predict low BMD fulfilled a mean of 10.8 items, with a range of 8-15, in 19 possible modified QUADAS items. Only 6 studies [15, 16, 19, 20, 25, 30] could be considered population based, meaning, including unselected men from the general population that would receive the test in a usual screening program. Most of the studies were relatively small (n=17) and it was unclear if the data regarding risk factors was collected blindly to DXA results and also if the DXA was interpreted blindly to the screening tool result. None of the selected studies reported information on uninterpretable test results. Further information on each QUADAS item is presented in Fig. 2.

Discussion

We identified 22 studies developing and/or evaluating the performance of clinical screening tools aimed to identify men with low BMD: 2 studies assessed the performance of 5 guidelines for DXA testing proposed from different entities; 5 studies developed and validated 5 new screening tools but of these only 2 were further validated in other populations; 12 studies evaluated the performance of OST, 4 assessed the performance of MORES and 3 estimated the performance of OSTA.

Over the last twenty years several tools have been developed aimed to identify individuals with low BMD. However, only ten years later, as seen from the present review, did these screening instruments start to be validated in men and since then there has been a growing interest in their role in this specific population. As far as we know, there is only one previous systematic review on this specific topic, dated from 2008 [6]. However, subsequently, an important number of new papers were published on this topic, justifying the discrepancy between the final number of studies included then and now (5 versus 22).

In general, when positive labelling from test results does not constitute a problem for the individual, screening instruments and their cut-offs should prefer to optimise sensitivity (i.e. minimizing the likelihood ratio of a negative test result) in the target population, which in this case is the general male adult population or the fraction that presents to primary health care. Likelihood ratios below 0.1 are considered to provide strong evidence to rule out the presence of disease [33]. The LR analysis performed in our review showed that negative likelihood ratios for OST and MORES were lower than for OSTA (0.00 versus 0.14) but also that, in one study, a negative LR of 1.00 was calculated for OST. At the elected cut-off by each OST study, negative LR ranged from 0.11 to 0.60, showing that most of these thresholds are labelling as healthy many individuals with low BMD.

Overall, tools screening for low BMD in men seem to have acceptable accuracy at some of the tested cut-offs. All the tools identified performed similarly across studies and the evidence is not sufficiently robust to determine which one should be adopted in clinical practice. However, as OST, OSTA and MORES have been validated more times, and include only few clinical variables, they seem to be more appropriate for this purpose, especially in the context of a routine clinical examination.

For application in White populations, it seems more suitable to select either OST or MORES, since OSTA was developed in Asians and validated only once in Whites and with low accuracy [19]. One of the studies assessing MORES had very low methodological quality. The 3 remaining articles, using a cut-off of MORES≥6, evidenced that the good discriminatory capacity showed in the development study diminished in a population based validation study [20] and was only slightly lower in an outpatient clinic population validation study [22]. OST had a wide range of values proposed as ideal cut-off (from 2 to 6) but the most often indicated was OST≤2. A similar validity decrease in real-life settings was observed as for MORES, with validation in population based studies generating lower prediction ability than those conducted in outpatient clinics. MORES is based on three clinical variables (age, weight and chronic obstructive pulmonary disease) and uses a weighted scale for score calculation. OST comprises only age and weight but implies more arithmetic than MORES.

We only identified one study assessing the accuracy OST in a specific disease (rheumatoid arthritis) [31]. This study showed that OST is not accurate in this population, as shown by moderate sensitivity and low specificity.

In this study we used robust methodology following the standards established by the PRISMA group for systematic reviews development, namely a detailed definition of inclusion and exclusion criteria and rigorous information extraction. However, in contrast with the recommendation from this group, for practical reasons we used only one database (Medline) for the search. Despite Medline being a widely diffused database, we cannot exclude the possibility of selection bias.

We also assessed the quality of studies based on a modified QUADAS check list. As suggested by the author, the QUADAS items should be adjusted for the specific topic of the review [9], which we did. The individual quality of studies included varied greatly making it difficult to summarize quantitatively the performance of the different screening osteoporosis tools developed to identify men with low BMD benefiting from DXA-testing. Also, in many cases, there was limited reporting of studies characteristics making formal quality assessment more difficult which is in itself a methodological limitation. This limitation has already been reported on previous systematic reviews of clinical decision rules for low BMD in women [8, 10, 34].

Conclusions

Overall, tools screening for low BMD in men seem to have acceptable accuracy but no tool performed consistently better than other. OST and MORES, for their simplicity, accuracy and replication, seem to be more adequate for routine clinical practice.

References

1. Ammann, P. and R. Rizzoli, *Bone strength and its determinants.* Osteoporos Int, 2003. **14 Suppl 3**: p. S13-8.

2. Bliuc, D., *Mortality risk associated with low-trauma osteoporotic fracture and subsequent fracture in men and women* JAMA, 2009. **4**(301 (5)): p. 513-521.

3. Strom, O., et al., *Osteoporosis: burden, health care provision and opportunities in the EU: a report prepared in collaboration with the International Osteoporosis Foundation (IOF) and the European Federation of Pharmaceutical Industry Associations (EFPIA).* Arch Osteoporos, 2011. **6**(1-2): p. 59-155.

4. McGinn, T.G., et al., *Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group.* Jama, 2000. **284**(1): p. 79-84.

5. Schwartz, E.N. and D.M. Steinberg, *Prescreening tools to determine who needs DXA.* Curr Osteoporos Rep, 2006. **4**(4): p. 148-52.

6. Liu, H., N.M. Paige, and C. Goldzweig, *Screening for Osteoporosis in Men: A Systematic Review for an American College of Physicians Guideline.* Annals of Internal Medicine, 2008. **148**(9): p. 685-701.

7. Liberati, A., et al., *The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration.* Ann Intern Med, 2009. **151**(4): p. W65-94.

8. Steurer, J., et al., *Clinical value of prognostic instruments to identify patients with an increased risk for osteoporotic fractures: systematic review.* PLoS One, 2011. **6**(5): p. e19994.

9. Whiting, P., et al., *The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews.* BMC Med Res Methodol, 2003. **3**: p. 25.

10. Rubin, K.H., et al., *Risk assessment tools to identify women with increased risk of osteoporotic fracture: complexity or simplicity? A systematic review.* J Bone Miner Res, 2013. **28**(8): p. 1701-17.

11. Verdijk, N.A., et al., *Validation of the Dutch guidelines for dual X-ray absorptiometry measurement.* Br J Gen Pract, 2009. **59**(561): p. 256-60.

12. Roig Vilaseca, D., et al., *[Sensitivity, specificity, positive and negative predictive values of the criteria for indicating a bone densitometry in the evaluation of medical techniques and research in Cataluna].* Reumatol Clin, 2011. **7**(3): p. 161-6.

13. Elliot, J., et al., *Historical assessment of risk factors in screening for osteopenia in a normal Caucasian population.* Aust NZ J Med, 1993. **23**: p. 458-462.

14. Zimering, M., et al., *Validation of a Novel Risk Estimation Tool for Predicting Low Bone Density in Caucasian and African American Men Veterans.* Journal of Clinical Densitometry, 2007. **10**: p. 289-297.

15. Shepherd, A., et al., *Development and Internal Validation of the Male Osteoporosis Risk Estimation Score.* Ann Fam Med 2007. **5**: p. 540-546.

16. Scholtissen, S., et al., *Assessment of determinants for osteoporosis in elderly men.* Osteoporos Int, 2009. **20**(7): p. 1157-66.

17. Kung, A.W., et al., *Development of a clinical assessment tool in identifying Asian men with low bone mineral density and comparison of its usefulness to quantitative bone ultrasound.* Osteoporos Int, 2005. **16**(7): p. 849-55.

18. Li-Yu, J.T., L.J. Llamado, and T.P. Torralba, *Validation of OSTA among Filipinos.* Osteoporos Int, 2005. **16**(12): p. 1789-93.

19. Machado, P., M. Coutinho, and J.A. da Silva, *Selecting men for bone densitometry: performance of osteoporosis risk assessment tools in Portuguese men.* Osteoporos Int, 2010. **21**(6): p. 977-83.

20. Shepherd, A.J., A.R. Cass, and L. Ray, *Determining risk of vertebral osteoporosis in men: validation of the male osteoporosis risk estimation score.* J Am Board Fam Med, 2010. **23**(2): p. 186-94.

21. Fransiska, Y., et al., *The male osteoporosis risk estimation score and the osteoporosis self-assessment screening tool for Indonesian men.* J Orthop Surg (Hong Kong), 2012. **20**(2): p. 205-8.

22. Cass, A. and A. Shepherd, *Validation of the Male Osteoporosis Risk Estimation Score (MORES) in a Primary Care Setting.* JABFM J, 2013. **26**: p. 436-444.

23. Pongchaiyakul, C. and E. Wanothayaroj, *Performance of the Khon Kaen Osteoporosis Study (KKOS) score for identifying osteoporosis in men.* J Med Assoc Thai, 2007. **90**(8): p. 1518-23.

24. Adler, R., M. Tran, and V. Petkov, *Performance of the Osteoporosis Self-assessment Screening Tool for osteoporosis in American men.* Mayo Clin Proc, 2003. **78**: p. 723–7.

25. Lynn, H., et al., *An osteoporosis screening tool for Chinese men.* Osteoporos Int 2005. **16**: p. 829–834.

26. Sinnott, B., S. Kukrejas, and E. Barengolts, *Utility of screening tools for the prediction of low bone mass in African American men.* Osteoporos Int 2006. **17**(5): p. 684-92.

27. Skedros, J.G., C.L. Sybrowsky, and G.J. Stoddard, *The osteoporosis self-assessment screening tool: a useful tool for the orthopaedic surgeon.* J Bone Joint Surg Am, 2007. **89**(4): p. 765-72.

28. Ghazi, M., et al., *Performance of the osteoporosis risk assessment tool in Moroccan men.* Clin Rheumatol, 2007. **26**(12): p. 2037-41.

29. Perez-Castrillon, J.L., et al., *OST risk index and calcaneus bone densitometry in osteoporosis diagnosis.* J Clin Densitom, 2007. **10**(4): p. 404-7.

30. Lynn, H., et al., *An evaluation of osteoporosis screening tools for the osteoporotic fractures in men (MrOS) study.* Osteoporos Int 2008. **19**: p. 1087–1092.

31. Richards, J.S., et al., *Dual-energy X-ray absorptiometry and evaluation of the osteoporosis self-assessment tool in men with rheumatoid arthritis.* J Clin Densitom, 2009. **12**(4): p. 434-40.

32. Richards, J., et al., *Validation of the Osteoporosis Self-Assessment Tool in US Male Veterans.* J Clin Densitom, 2013 **12**: p. 1-6.

33. Jaeschke, R., G. Guyatt, and J. Lijmer, *Diagnostic Tests*, ed. AmericanMedicalAssociation. 2002.

34. Rud, B., et al., *Performance of the Osteoporosis Self-Assessment Tool in ruling out low bone mineral density in postmenopausal women: a systematic review.* Osteoporos Int, 2007. **18**(9): p. 1177-87.

Table 1. Characteristics of Clinical Screening Tools predicting T-scores ≤-2.5

Author, Year [Reference]	Local	Sample characteristics	Men, *n*	Age (mean or range)	Tool	DXA-Equipment (Model)	Reference Group (age, sex, population, race)	Anatomical site for low BMD diagnosis	Sensitivity	Specificity	Cut-Off	AUC	LR+	LR-
Adler et al., 2003 [24]	USA/ Canada	Pulmonary and rheumatology clinics	181	64.3	OST	Hologic (QDR 4500)	Manufacture reference (LS); NHANES (TH/FN); other characteristics not mentioned	Lumbar spine, total hip or femoral neck	93% 95% 83% 82% 74%	66% 60% 80% 74% 80%	OST<3 OST<2 OST<1 OST<3 (W) OST<3 (AA)	0.836 0.848 (W) 0.800 (AA)	2.74 2.38 4.15 3.15 3.70	0.11 0.08 0.21 0.24 0.32
Lynn et al., 2005 [25]	China	Population based	1970	73.0	OST	Hologic (QDR 4500)	Young adult men; Chinese	Lumbar spine, total hip or femoral neck	NA	NA	NA	0.774	NA	NA
Kung et al., 2005 [17]	China	Population based	356	65.0 and 64.0	OSTA	Hologic (QDR 2000)	Young adult men; Chinese	Lumbar spine, total hip or femoral neck	83% 72% 71%	67% 65% 68%	OSTA≤-1 (FN) OSTA≤-1 (LS) OSTA≤-1 (AS)	0.850 0.790 0.780	2.52 2.06 2.22	0.25 0.43 0.43
Li-Yu et al., 2005 [18]	Philippines	Patients referred to DXA	132	62.9	OSTA	GE-Lunar (DPX IQ)	Young adult men; Chinese	Femoral neck	91%	66%	OSTA≤-1	0.848	2.68	0.14
Sinnott et al., 2006 [26]	USA	Afro-American from outpatient general clinic	128	63.8	OST	GE-Lunar (NA)	Young women, manufacture population, White (LS); young male, manufacture population, Caucasian (TH/FN)	Total hip and femoral neck	89% 89% 89%	54% 64% 74%	OST<4 OST<3 OST<2	0.890	1.93 2.47 3.42	0.20 0.17 0.15

Table 1. Continued

Author, Year [Reference]	Local	Sample characteristics	Men, *n*	Age (mean or range)	Tool	DXA-Equipment (Model)	Reference Group (age, sex, population, race)	Anatomical site for low BMD diagnosis	Sensitivity	Specificity	Cut-Off	AUC	LR+	LR-
Skedros et al., 2007 [27]	USA	Orthopaedic clinic	158	67.5	OST	GE-Lunar (Prodigy)	NA	Lumbar spine, total hip or femoral neck	85%	64%	OST<2	0.760	2.36	0.23
Ghazi et al., 2007 [28]	Morocco	Patients referred to DXA	229	62.3	OST	GE-Lunar (Prodigy)	Young women, Moroccan	Lumbar spine, total hip or femoral neck	64% 88% 64% 52% 75% 54% 36% 63% 40%	60% 58% 60% 71% 70% 73% 80% 79% 81%	OST<2 (LS) OST<2 (TH) OST<2 (AS) OST<1 (LS) OST<1 (TH) OST<1 (AS) OST<0 (LS) OST<0 (TH) OST<0 (AS)	0.660 (LS) 0.787 (TH) 0.667 (AS)	1.60 2.10 1.60 1.79 2.50 2.00 1.80 3.00 2.11	0.60 0.21 0.60 0.68 0.36 0.63 0.80 0.47 0.74
Castrillón et al., 2007 [29]	Spain	Outpatient general clinic (with suspected osteoporosis)	67	47.0	OST	GE-Lunar (DPX L) and Hologic (QDR 4500)	Young (sex not mentioned); manufacture population	Lumbar spine, total hip or femoral neck	NA	NA	OST<3	0.623 (non significant)	NA	NA
Pongchaiyakul et al., 2007 [23]	Taiwan	Population based	230	63.4	KKOS	GE-Lunar (DPX IQ)	Young men, Thai population	Lumbar spine or femoral neck	94% 72% 73%	70% 73% 73%	KKOS<-1 (LS) KKOS<-1 (FN) KKOS<-1 (AS)	NA	3.13 2.67 2.70	0.09 0.38 0.37

Table 1. Continued

Author, Year [Reference]	Local	Sample characteristics	Men, *n*	Age (mean or range)	Tool	DXA-Equipment (Model)	Reference Group (age, sex, population, race)	Anatomical site for low BMD diagnosis	Sensitivity	Specificity	Cut-Off	AUC	LR+	LR-
Shepherd et al., 2007 [15]	USA	Population based	1498	64.2	MORES	Hologic (NA)	Young male, NHANES population, specific for race	Total hip	95%	61%	MORES≥6	0.842	2.44	0.08
Zimering et al., 2007 [14]	USA	Endocrinology, rheumatology and general clinics	197 (C) 134 (AA)	68.4 (C) 60.9 (AA)	Mscore Mscore$_{age\text{-}weight}$; OST	Hologic (QDR 4500)	Young male, NHANES population, specific for race	Femoral neck	88%	49%	Mscore>8 (W)	0.840	1.73	0.24
									88%	57%	Mscore>9 (W)		2.05	0.21
									85%	50%	Mscore$_{age\text{-}weight}$>8 (W)	0.810	1.70	0.30
									85%	58%	Mscore$_{age\text{-}weight}$>9 (W)		2.02	0.26
									75%	65%	Mscore$_{age\text{-}weight}$>10 (W)		2.14	0.38
									100%	69%	Mscore$_{age\text{-}weight}$>8 (AA)	0.990	3.23	0.00
									100%	73%	Mscore$_{age\text{-}weight}$>9 (AA)		3.70	0.00
									100%	77%	Mscore$_{age\text{-}weight}$>10 (AA)		4.35	0.00
									75%	68%	OST<2 (W)	0.810	2.34	0.37
									75%	59%	OST<3 (W)		1.83	0.42
									85%	51%	OST<4 (W)		1.73	0.29
									100%	83%	OST<2 (AA)	0.990	5.88	0.00
									100%	76%	OST<3 (AA)		4.17	0.00
									100%	72%	OST<4 (AA)		3.57	0.00

Table 1. Continued

Author, Year [Reference]	Local	Sample characteristics	Men, *n*	Age (mean or range)	Tool	DXA-Equipment (Model)	Reference Group (age, sex, population, race)	Anatomical site for low BMD diagnosis	Sensitivity	Specificity	Cut-Off	AUC	LR+	LR-
Lynn et al., 2008 [30]	USA and China	Population based	4658 Caucasian 1914 Chinese	65-?	OST	Hologic (QDR 4500)	Whites: young men, NHANES population, white (FN) and young men, manufacture population (LS) Chinese: young men, Chinese population;	Lumbar spine, total hip or femoral neck	79% 88%	49% 36%	OST<1 OST<2	Whites: 0.662 (LS) 0.823 (TH) 0.740 (FN) 0.714 (AS)	1.55 1.38	0.43 0.33
									82% 91%	56% 36%	OST<1 OST<2	Chinese: 0.717 (LS) 0.855 (TH) 0.849 (FN) 0.759 (AS)	1.86 1.42	0.32 0.25
Scholtissen et al., 2009 [16]	Belgium and France	Population Based	502	69.3	Age + BMI + Family history of fractures + Previous fracture	Hologic (QDR 4500) and GE-Lunar (Prodigy)	Young men, local population	Lumbar spine, total hip or femoral neck	NA	NA	NA	0.712	NA	NA

Table 1. Continued

Author, Year [Reference]	Local	Sample characteristics	Men, *n*	Age (mean or range)	Tool	DXA-Equipment (Model)	Reference Group (age, sex, population, race)	Anatomical site for low BMD diagnosis	Sensitivity	Specificity	Cut-Off	AUC	LR+	LR-
Verdijket et al., 2009 [11]	Netherlands	Outpatient general clinic	64	71.7	Dutch case-find instrument	Hologic (QDR 4500)	FN: young (sex not mentioned), NHANES population; LS: young (sex not mentioned), manufacture population	Lumbar spine and femoral neck	11%	92.7%	≥4	NA	1.51	0.96
Machado et al., 2010 [19]	Portugal	Population based	202	63.8	OSTA and OST	Hologic (QDR 4500)	FN: young (sex not mentioned), NHANES population; LS: young male, manufacture population	Lumbar spine, total hip or femoral neck	47% 62% 77% 85% 38% 56% 74% 77%	73% 64% 50% 33% 82% 68% 58% 43%	OST<1 OST<2 OST<3 OST<4 OSTA<1 OSTA<2 OSTA<3 OSTA<4	NA	1.74 1.72 1.54 1.27 2.11 1.75 1.76 1.35	0.73 0.59 0.46 0.45 0.76 0.65 0.45 0.53
Richards et al., 2009 [31]	USA	Men with rheumatoid arthritis	282	66.1	OST	Hologic (NA) and Lunar (NA)	Young male, NHANES population, specific for race	Femoral neck and total hip	6% 18% 34% 40% 58% 68% 78% 84%	94% 92% 80% 71% 64% 53% 45% 34%	OST≤-2 OST≤-1 OST≤0 OST≤1 OST≤2 OST≤3 OST≤4 OST≤5	NA	1.00 2.25 1.70 1.38 1.61 1.45 1.42 1.27	1.00 0.89 0.83 0.85 0.66 0.60 0.49 0.47

Table 1. Continued

Author, Year [Reference]	Local	Sample characteristics	Men, *n*	Age (mean or range)	Tool	DXA-Equipment (Model)	Reference Group (age, sex, population, race)	Anatomical site for low BMD diagnosis	Sensitivity	Specificity	Cut-Off	AUC	LR+	LR-
Shepherd et al., 2010 [20]	USA	Population based	2944	63.0	MORES	Hologic (QDR 4500)	Young male, population not mentioned	Lumbar spine	58% 51% 76% 60% 90%	65% 67% 62% 56% 50%	MORES>6	0.657 0.653 (W) 0.786 (AA) 0.601 (MA) 0.648 (others)	1.29 1.55 2.00 1.36 1.8	0.65 0.73 0.39 0.71 0.20
								Lumbar spine, total hip or femoral neck	66% 60% 79% 71% 95%	68% 69% 63% 59% 55%	MORES>6	0.728 0.721 (W) 0.781 (AA) 0.703 (MA) 0.675 (others)	2.06 1.94 2.14 1.73 2.11	0.50 0.58 0.33 0.49 0.09
Vilaseca et al., 2011 [12]	Spain	Outpatient rheumatology and surgery clinics	49 + 52	56.5 and 72.7	Guidelines AETIM, NOF 2010, WHO, WHO 2003	Hologic (QDR 1000)	Sex or age not mentioned; Spanish population	Lumbar spine and femoral neck	60% 67%	91% 67%	AETIM	NA	6.67 2.06	0.44 0.49
									100% 100%	32% 8%	NOF	NA	1.47 1.09	0.00 0.00
									100% 83%	33% 25%	WHO	NA	1.49 1.11	0.00 0.68
									80% 50%	56% 61%	WHO 2003	NA	1.82 1.28	0.36 0.82
Fransiska et al., 2012 [21]	Indonesia	NA	113	71.0	MORES; OST	NA	NA	NA	100%	7%	MORES≥6	0.535	1.08	0.00
									74%	41%	OST≤2	0.574	1.25	0.63

Table 1. Continued

Author, Year [Reference]	Local	Sample characteristics	Men, *n*	Age (mean or range)	Tool	DXA-Equipment (Model)	Reference Group (age, sex, population, race)	Anatomical site for low BMD diagnosis	Sensitivity	Specificity	Cut-Off	AUC	LR+	LR-
Richards et al., 2013 [32]	USA	Outpatient general clinic	518	66.0	OST	Hologic (NA) and Lunar (NA)	young male, NHANES population, specific for race	Lumbar spine, total hip or femoral neck	7% 42% 62% 75% 86% 90%	98% 85% 65% 41% 32% 23%	OST≤-3 OST≤-0 OST≤3 OST≤5 OST≤6 OST≤8	0.720 (W)	3.50 2.80 1.77 1.27 1.26 1.17	0.95 0.68 0.58 0.61 0.44 0.43
									10% 14% 45% NA 70% 83%	99% 99% 69% NA 36% 30%	OST≤-3 OST≤-0 OST≤3 OST≤5 OST≤6 OST≤8	0.580 (AA)	10 14 1.45 NA 1.04 1.19	0.91 0.87 0.80 NA 0.83 0.57
Cass et al., 2013 [22]	USA	Outpatient general clinic	346	70.2	MORES	Hologic (QDR 4500) and GE-Lunar (iDXA)	young women, NHANES population	Femoral neck or total hip	80%	70%	MORES ≥6	0.82	2.67	0.29

DXA, Dual-energy X-ray absorptiometry; BMD, bone mineral density; AUC, Area under the curve; LR+, positive likelihood ratio; LR-, negative likelihood ratio; OST, Osteoporosis screening tool; W, Whites; AA, African-American; NA, Non-available; OSTA, osteoporosis screening tool for Asians; FN, femoral neck; LS, lumbar spine; AS, any site; USA, United States of America; TH, total hip; KKOS, Khon Kaen osteoporosis study; MORES, male osteoporosis risk estimation score; NHANES, National health and nutrition examination study; MA, Mexican American.

Table 2. Characteristics of Clinical Screening Tools predicting T-scores ≤-2.0

Author, Year [Reference]	Local	Sample characteristics	Men, *n*	Age (mean or range)	Tool	DXA-Equipment (Model)	Reference Group (age, sex, population, race)	Anatomical site for low BMD diagnosis	Sensitivity	Specificity	Cut-Off	AUC	LR+	LR-
Adler et al., 2003 [24]	USA/ Canada	Pulmonary and rheumatology clinics	181	64.3	OST	Hologic (QDR 4500)	Manufacture reference (LS); NHANES (TH/FN); other characteristics not mentioned	Lumbar spine, total hip or femoral neck	74% 69% 62%	72% 82% 89%	OST<3 OST<2 OST<1	0.815	2.64 3.83 5.64	0.36 0.38 0.43
Li-Yu et al., 2005 [18]	Philippines	Patients referred to DXA	132	62.9	OSTA	GE-Lunar (DPX IQ)	Young adult men; Chinese	Femoral neck	72%	69%	OSTA≤-1	0.754	2.32	0.41
Sinnott et al., 2006 [26]	USA	Afro-American from outpatient general clinic	128	63.8	OST	GE-Lunar (NA)	Young women, manufacture population, Whites (LS); young male, manufacture population, Whites (TH/FN)	Total hip and femoral neck	83% 78% 71%	57% 68% 76%	OST<4 OST<3 OST<2	0.830	1.93 2.43 2.96	0.30 0.32 0.38
Skedros et al., 2007 [27]	USA	Orthopaedic clinic	158	67.5	OST	GE-Lunar (Prodigy)	NA	Lumbar spine, total hip or femoral neck	76%	67%	OST<2	0.750	2.30	0.36

Table 2. (continued)

Author, Year [Reference]	Local	Sample characteristics	Men, *n*	Age (mean or range)	Tool	DXA-Equipment (Model)	Reference Group (age, sex, population, race)	Anatomical site for low BMD diagnosis	Sensitivity	Specificity	Cut-Off	AUC	LR+	LR-
Lynn et al., 2008 [30]	USA and China	Population based	4658 Caucasian + 1914 Chinese	65-?	OST	Hologic (QDR 4500)	Chinese: young men, Chinese population; Whites: young men, NHANES population, Whites (FN) and young men, manufacture population (LS)	Lumbar spine, total hip or femoral neck	NA	NA	NA	**Caucasian:** 0.648 (LS) 0.745 (TH) 0.710 (FN) 0.695 (AS)	NA	NA
									NA	NA	NA	**Chinese:** 0.694 (LS) 0.791 (TH) 0.758 (FN) 0.724 (AS)	NA	NA

DXA, Dual-energy X-ray absorptiometry; BMD, bone mineral density; AUC, Area under the curve; LR+, positive likelihood ratio; LR-, negative likelihood ratio; USA, United States of America; OST, Osteoporosis screening tool; LS, lumbar spine; NHANES, National health and nutrition examination study; TH, total hip; FN, femoral neck; OSTA, osteoporosis screening tool for Asians; NA, Non-available; AS, any site.

Table 3. Characteristics of Clinical Screening Tools predicting T-scores ≤ -1.0

Author, Year [Reference]	Local	Sample characteristics	Men, *n*	Age (mean or range)	Tool	DXA-Equipment (Model)	Reference Group (age, sex, population, race)	Anatomical site for low BMD diagnosis	Sensitivity	Specificity	Cut-Off	AUC	LR+	LR-
Elliot et al., 1993 [13]	New Zealand	Population based	126	20-82	Weight + inactivity	GE-Lunar (DPX 4)	BMD in the lower third of the age matched normal range	Lumbar spine	83%	46%	0.054	NA	1.54	0.37
					Age + Weight + family history + inactivity + weekly calcium			Femoral neck	87%	45%	0.040	NA	1.58	0.29
Vilaseca et al., 2011 [12]	Spain	Outpatient rheumatology and surgery clinics	49 + 52	56.5 and 72.7	Guidelines from AETIM, NOF 2010, WHO and WHO 2003	Hologic (QDR 1000)	Sex or age not mentioned; Spanish population	Lumbar spine and femoral neck	18% 43%	92% 75%	AETIM	NA	2.25 1.72	0.89 0.76
									79% 97%	39% 17%	NOF	NA	1.30 1.17	0.54 0.18
									76% 83%	39% 42%	WHO	NA	1.25 1.43	0.62 0.40
									68% 47%	68% 75%	WHO 2003	NA	2.13 1.88	0.47 0.71

Table 3. Continued

Author, Year (Reference)	Local	Sample characteristics	Men, *n*	Age (mean or range)	Tool	DXA-Equipment (Model)	Reference Group (age, sex, population, race)	Anatomical site for low BMD diagnosis	Sensitivity	Specificity	Cut-Off	AUC	LR+	LR-
Richards et al., 2009 [31]	USA	Men with rheumatoid arthritis	282	66.1	Validation of OST	Hologic (not mentioned) and Lunar (not mentioned)	Young male, NHANES population, specific for race	Femoral neck and total hip	7%	98%	OST≤-2	NA	3.50	0.94
									13%	99%	OST≤-1		13	0.88
									27%	90%	OST≤0		2.7	0.81
									35%	79%	OST≤1		1.67	0.82
									46%	74%	OST≤2		1.77	0.73
									57%	65%	OST≤3		1.63	0.66
									64%	54%	OST≤4		1.39	0.67
									73%	42%	OST≤5		1.26	0.64

DXA, Dual-energy X-ray absorptiometry; BMD, bone mineral density; AUC, Area under the curve; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NA, Non-available; AETIM, *Agencia de Evaluación de Tecnologías e Investigación Médicas*; NOF, National Osteoporosis Foundation; WHO, World Health Organization; USA, United States of America; OST, Osteoporosis screening tool; NHANES, National health and nutrition examination study.

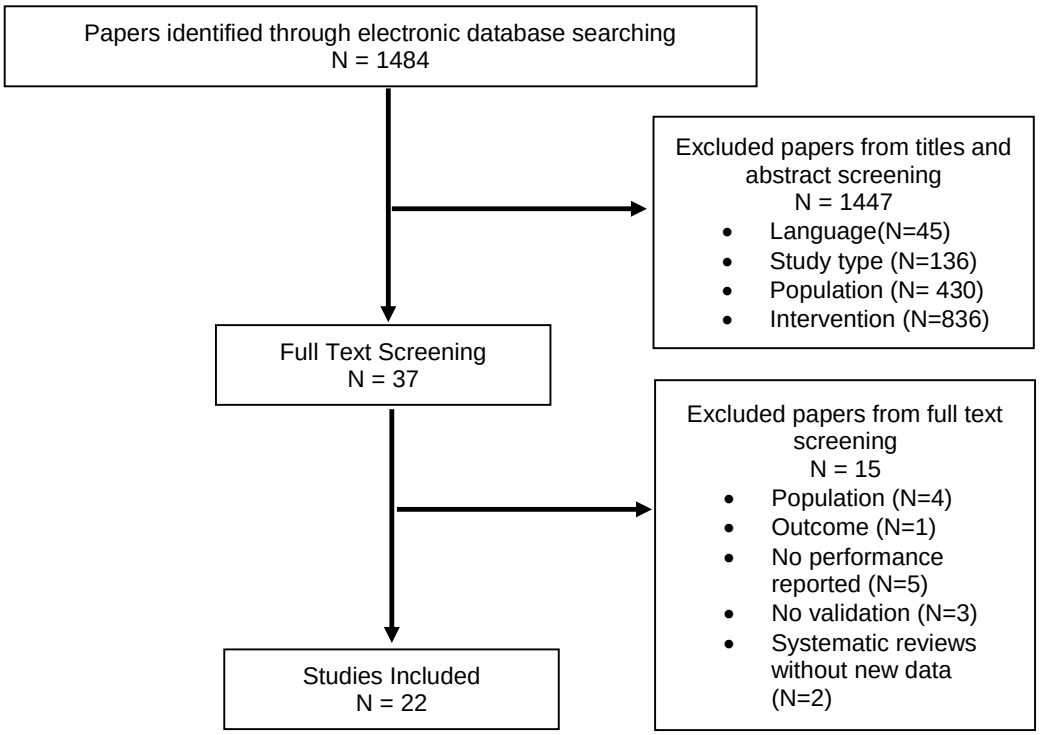

Figure 1. Flowchart of included studies.

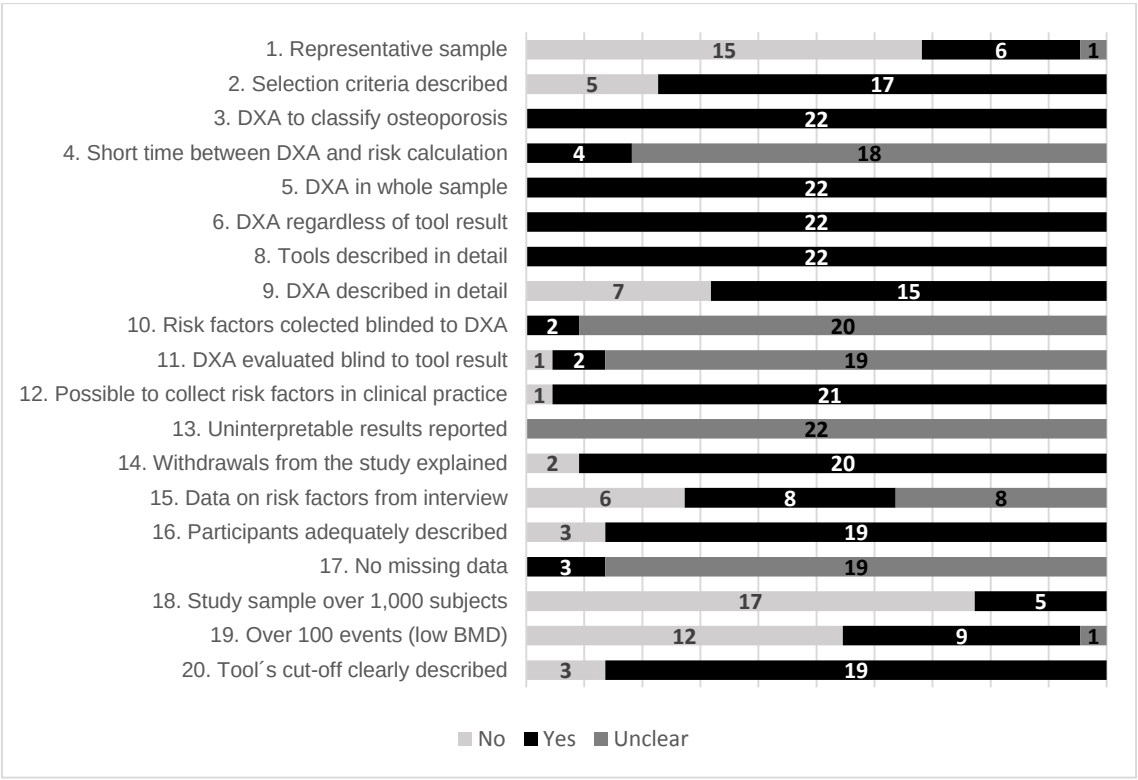

Figure 2. Methodological quality of 22 included studies according to a modified QUADAS checklist. QUADAS, Quality assessment tool for diagnostic accuracy studies; BMD, Bone mineral density; DXA, Dual-energy X-Ray absorptiometry.

CHAPTER II

Upper arm circumference measurement improves screening for low bone mineral density in men

Romana Vieira, Milton Severo, Carla Lopes, Raquel Lucas

Abstract

Objectives: To develop and validate a simple clinical screening tool (EPIPOST) able to identify men with higher probability of having low bone mineral density (BMD) who may benefit from dual-energy X-ray absorptiometry (DXA) testing and to compare its discriminatory ability with two other osteoporosis screening tools in men (OST and MORES), after calibration for our population. **Methods:** As part of EPIPorto population-based study among adults, 147 men aged between 40 and 65 years were assessed. Age, height, weight, body mass index and several body circumferences were recorded by trained observers. DXA whole-body scans were performed for BMD assessment. For the calibration of OST and MORES, new regression parameters were estimated for each risk factor included accounting for their prevalence and also for the prevalence of low BMD in our population. For EPIPOST development, the different anthropometric variables were tested using logistic regression models to predict low BMD. EPIPOST validation was done by the leave-one-out cross-validation method. The overall fit and discriminatory capacity of the different models were assessed by direct comparison of the observed and expected prevalences of low BMD by quartiles of each score, Hosmer-Lemeshow "goodness-of-fit" test and area under the receiver operating characteristic (ROC) curve. Finally, likelihood ratios (LR) were calculated to select the ideal cut-off for each model. **Results:** Calibration maintained the discriminatory capacity of OST and MORES (AUC of 0.73 and 0.75, respectively) and improved the fit. The EPIPOST included only upper arm circumference and showed better discriminatory capacity (AUC 0.76). For predicting low BMD, OST≤2 had a sensitivity of 100% and a specificity of 8.2%; MORES>-2 had a sensitivity of 93.9% and a specificity of 30.6%; EPIPOST>-2 had a sensitivity of 98.0% and a specificity of 18.6%. The LR analysis showed that EPIPOST had higher discriminative ability across different risk levels (LR range of 0.1 to 18.4, compared to 0.0 to 2.4 with OST and 0.2 to 2.8 with MORES). **Conclusion:** Calibration of OST and MORES improved the fitting of both models to our population data and maintained their discriminatory ability to identify men with low BMD. The newly developed tool, EPIPOST, is easier to execute in clinical practice and performed similarly to OST and MORES in terms of overall accuracy while showing a wider range of discriminatory ability.

Keywords: Men, Bone Density, Clinical Prediction Rule, Sensitivity and Specificity, Screening.

Introduction

Lifetime risk of osteoporotic fracture is higher in women [1]. However, evidence shows that case fatality rate after osteoporotic fracture is higher in men [2-4]: the relative risk of dying within 1 year after hip fracture versus controls has been estimated in 4.2 in men and 3.3 in women below 75 years of age [5]. Nevertheless, comparatively little attention to the diagnosis or treatment of male osteoporosis has been given [6], even though the early recognition of osteoporosis in men may also have key importance for the disease burden in terms of years of life lost.

The World Health Organization (WHO) recommends the identification of individuals eligible for antiresorptive treatment based on a comprehensive fracture risk assessment, namely with a dedicated tool (FRAX™). Nevertheless, bone mineral density (BMD) assessment by dual-energy X-ray absorptiometry (DXA) remains the gold standard method for the diagnosis of osteoporosis [7] and BMD is the strongest objectively measured determinant of bone strength accounting for 60-70% of its variation [8]. Because of that, BMD determination is still a cornerstone of clinical practice, not only to assess fracture risk but also for selecting patients for treatment and to evaluate their subsequent response [9].

Even though DXA equipment is available in many settings, it was estimated that, to detect one previously undiagnosed case of osteoporosis, the number needed to screen by BMD testing is high in the population with no clinical risk factors and even higher in men than women (6 women aged 65 years old or more, 13 men aged 65 years old or more and 10 men aged 70 years old or more) [10]. This calls for a targeted case finding strategy rather than population-based screening in order to maximize yield and cost-effectiveness [9]. Simple tools able to identify men at higher risk of osteoporosis using only easily obtained clinical data facilitate case ascertainment namely in primary healthcare settings.

Several clinical tools have been developed and validated in women [11-15] based on variables like age and anthropometric measures [16-19]. The Osteoporosis Self-assessment Tool (OST) was originally developed in women [11] but has also been validated in men [20-22]. Other tools have been specifically developed and validated for

men [23-27]. Because of their strong association with bone fragility, age and weight [16] have been included in most of these parsimonious models. However, risk factors for bone fragility are likely to operate differently across genders: with ageing, women tend to experience accelerated bone loss after menopause whereas men have a more gradual one [28]; also weight probably translates different risks because of sexual dimorphism related to body composition, with the ratio of lean to fat being much greater in males than females [29].

We hypothesised that the inclusion of easily accessible anthropometric measures that reflect sexual dimorphism of body composition would have similar accuracy to age and weight in predicting low bone mineral density in men, when compared to previously developed tools.

Our objective was to develop and validate a simple clinical screening tool (EPIPOST) able to identify men with higher probability of having low BMD who may benefit from DXA-scan testing, based on variables easily obtained, taking into account sexual dimorphism. We also compared the performance (discrimination) of our tool with the ones of the two most validated osteoporosis screening tools in men [OST and MORES (male osteoporosis risk estimation score)], after calibration for our population.

Methods

Participants

In 2010/2011, we evaluated 147 men aged between 40 and 65 years old. They are a subsample selected randomly and stratified by sex and age within the EPIPorto cohort which is a sample of urban dwellers in Porto. Briefly, participants were selected between 1999 and 2003, using random digit dialling of landline telephones to select households. The vast majority of houses (>95%) had a landline telephone at the time of this procedure. We used a table of random numbers to define the last four digits that are specific to individual houses, assuming the local prefix codes to limit the universe to the city of Porto. Within each household, we selected a permanent resident, aged 18 years or more, using simple random sampling. We considered a refusal if the person explicitly said that he/she did not want to participate and refusals were not substituted within the same household. The proportion of participation was 70%, as previously described [30].

Data collection

Trained interviewers administered a structured questionnaire comprising questions on sociodemographic, clinical and behavioural characteristics.

Anthropometric measurements were performed, according to standard procedures, after an overnight fast, with the participant wearing light clothing and no footwear [31]. Body weight was measured to the nearest 0.1 kg using a digital scale, and height was measured to the nearest centimetre in the standing position using a wall stadiometer. Body mass index (BMI) was calculated as weight (kg) divided by squared height (m^2). Waist circumference was measured midway between the lower limit of the rib cage and the iliac crest. Abdominal circumference was measured at the umbilicus level with the individual at full expiration. Measurements were taken to the nearest 0.1 cm**.** Hip circumference was measured in orthostatic position considering the highest value at the gluteus level to the nearest 0.1 cm. Right mid-thigh circumference was measured to the nearest 0.1 cm with the individual in sitting position, with the leg at 90º flexion, at the midpoint between the inguinal line and the superior patellar border. Relaxed upper arm circumference was measured at the non-dominant arm at the mid-point between scapula acromial apophysis and the radius head. The arm was relaxed along the body. The tape was tightly positioned without pressure, parallel to the floor. Measurements were taken to the nearest 0.1 cm. Flexed upper arm circumference was measured at the non-

dominant arm in the highest point. The tape was tightly positioned without pressure, parallel to the floor. Measurements were taken to the nearest 0.1 cm.

DXA whole-body scans were performed for BMD assessment (QDR 4500A, Hologic, Bedford, MA). A single DXA operator conducted all scans. We used T-scores as a relative measure of BMD as recommended by WHO. T-score is a value for BMD expressed as the number of standard deviations (SD) by which an individual measurement differs from the mean value for healthy adult female population aged 30 years [32]. Definition of osteoporosis was refined with the femoral neck being proposed as the standard measurement site and the reference population being the mean and SD values in young women obtained from the NHANES III study [33, 34]. Participants were classified as normal (T-score≥-1.0), osteopenic (-1.0>T-score>-2.5) and osteoporotic (T-score≤ -2.5) [35]. Finally, we dichotomized T-score value at -1.0 with participants with a T-score<-1.0 being classified as having low bone mineral density (osteopenia + osteoporosis).

Statistical analysis

Student's t-test was computed to compare the distributions of continuous variables between two independent samples. All p-values were two sided, and the significance level was set at 5%.

For the calibration of OST and MORES, new regression parameters were estimated for each risk factor included accounting for their prevalence and also for the prevalence of low BMD in our population, thereby optimizing the fit of the resulting model to the observed data. The calibrated parameters are presented in Appendix 1. For the evaluation of fit we performed a direct comparison of the observed and expected prevalences of low BMD by quartiles of each score and also applied the Hosmer-Lemeshow "goodness-of-fit" test before and after calibration for each model.

For the new tool (EPIPOST) development we first performed a bivariate analysis for the association of low bone mineral density and the candidate predictive variables. Then we ran logistic regression models using the forward method to predict low bone mineral density. The final model was selected taking into account the clinical simplicity, face validity, discriminatory ability based on area under the receiver operating characteristic (ROC) curve, and overall fit (Hosmer-Lemeshow "goodness-of-fit" test).

Areas under the ROC curves (AUC) were calculated to assess the discriminatory capacity of the calibrated and new tools to differentiate between individuals with and without low BMD. DXA-derived T-score categories (low BMD if T-score<-1.0 and normal BMD if T-score≥-1.0) were used as the state variable for this. Finally, likelihood ratios (LR) were calculated to select the ideal cut-off for each model.

Validation was done by the leave-one-out cross-validation method. Briefly, the equation was trained on 147 minus 1 individuals, and the trained equation was then used to test the individual that had been left out. This process was repeated until every individual in the dataset had been used once as an un-seen test individual. The agreement estimated by leave-one-out cross-validation was then compared with the one estimated using the whole sample to evaluate possible over-fitting.

Statistical analysis was performed using IBM SPSS Statistics 21 and RStudio statistical software.

Results

A total of 147 men were included in the current analysis with a mean age of 54 years. Characteristics of the study participants are listed in Table 1 as well as their stratification by BMD category: participants with low BMD were lighter and had smaller overall circumferences than those with normal BMD.

The prevalence of osteoporosis was 0.7% (n=1) at the femoral neck and 13.6% (n=20) at the lumbar spine. Considering the presence of T-score≤-2.5 at least in one of the two sites, the overall prevalence of osteoporosis was 14.2% (n=21). Low BMD at the femoral neck was present in 49 participants (33.3%).

Weight, BMI and all circumferences were significantly associated with low BMD in bivariate analysis (Table 2). When the significant variables were entered into logistic regression model with forward selection method, the final model retained only relaxed upper arm circumference for low BMD prediction. In fact, when BMI and circumferences were adjusted for relaxed upper arm circumference, their statistical significance observed in crude analysis was lost. Therefore, the final estimated equation was EPIPOST ~ 11.6 - 0.39 * (relaxed upper arm circumference).

As shown in Figure1, calibration of OST and MORES enhanced the similarity between the observed prevalence of low BMD in our population and that predicted by the models in each score quartiles. Although OST and MORES scores relate differently to BMD (low BMD is associated with low OST score and high MORES score), to simplify the interpretation of the results, we inverted the OST equation so that higher OST scores relate also to low BMD. The fit of both screening tools tested by the Hosmer-Lemeshow goodness-of-fit test improved after calibration despite maintenance of overall accuracy as measured by the AUC value (Table 3). Mean (SD) OST score was 2.49 (1.51) and after calibration was -0.90 (1.10). Mean (SD) MORES score was 3.49 (3.04) and after calibration was 0.82 (0.98). Estimated parameters after calibration for both OST and MORES are presented in Appendix 1.

In terms of global accuracy comparison, the two calibrated models and the newly developed one had similar results. However, even though only marginally, EPIPOST had the highest AUC (0.76). The likelihood ratios (LR) analysis (Table 4) showed that for $OST_{calibrated}$ the ideal cut-off to rule out the presence of low BMD was 2, meaning that individuals scoring ≤2 should be referred to DXA-scan testing. For both $MORES_{calibrated}$

and EPIPOST the equivalent cut-off was -2, meaning that individuals scoring >-2 should undergo DXA-scan testing. However, the LR associated with these cut-offs were not the same: OST had the cut-off with lowest LR (0.0) followed by EPIPOST (0.1) and finally MORES (LR 0.2). Moreover, the range of LR values obtained with EPIPOST showed that our tool had higher discriminative ability across different risk levels (0.1 to 18.4, compared to 0.0 to 2.4 with OST and 0.2 to 2.8 with MORES).

An OST≤2 had a sensitivity of 100% and a specificity of 8.2% to predict low BMD. A MORES>-2 had a sensitivity of 93.9% and a specificity of 30.6%. An EPIPOST>-2 had a sensitivity of 98.0% and a specificity of 18.6%. The EPIPOST positive LR for that cut-off was 1.20 and the negative LR was 0.11.

The leave-one-out cross-validation showed that the agreement between the observed and predicted values was 73.3% in the training sample and 71.2% across validation samples. Chance-corrected agreement, measured by Cohen's kappa coefficient, was 0.35 and 0.28, respectively.

Discussion

We showed that calibration improved the OST and MORES fitting to our population and that individuals scoring OST≤2 and MORES>-2 should undergo DXA-scan testing. OST proved to have higher sensitivity than MORES (100% versus 93.9%) despite lower specificity (8.2% versus 30.6%) in identifying men with low BMD that should be evaluated with DXA testing. We also developed a new low BMD screening tool, EPIPOST, specifically for male sex. This new tool had even better discriminatory ability than OST and MORES (AUC 0.76 versus 0.73 and 0.75, respectively) and, at scores greater than -2, improved sensitivity regarding to MORES (98.0% versus 93.9%) and better specificity than OST (18.6% versus 8.2%).

As osteoporosis remains asymptomatic until fracture occurs, bone mass assessment and early low bone mass diagnosis is a cornerstone of fracture prevention. Although clinical assessment tools are not meant to diagnose osteoporosis they do assist clinicians in identifying asymptomatic individuals likely to have low bone mass and, thus, they are an important part of population approaches to the prevention of osteoporotic fractures. Earlier studies on osteoporosis screening tools focused their attention on older men with higher risk of osteoporosis. However, it is precisely in men over the age of 70 that immediate DXA screening proves to be more cost-effective, as supported by clinical guidelines, thereby limiting the need for prior clinical decision rules. Our objective was on the one hand to determine the clinical utility of clinical decision rules in a younger range of male subjects (40 to 65) and, on the other hand, to assess their capacity of identifying earlier stages of decreased bone mass when preventive measures may be more effectively applied.

We calibrated and evaluated the performance of the two most validated screening tools for the detection of low bone mineral density in men (OST and MORES). Risk prediction scores are developed and validated in target populations with a certain risk level. Independently of its performance in terms of overall accuracy, the discriminative power of each score in predicting the status of each individual in different populations may be increased through its calibration to the target population of interest. Regarding MORES, the coefficient for chronic obstructive pulmonary disease (COPD) was not calibrated as the prevalence of this condition in our population was very low (n=3). Also, since MORES includes age classes, with the highest risk being associated to age≥75 years, the fact that our population is younger than 65 years old probably affected MORES discriminatory

capacity and did not take advantage of the its whole spectrum of predictive ability. Still, calibration of both OST and MORES permitted the improvement of the agreement between observed and estimated prevalences of low BMD across scores' quartiles. As far as we know, no other previous study has been conducted to accomplish this objective.

Despite the achievement of a better discriminatory performance after calibration of these two models, we still found pertinent to develop a model that took into consideration sexual body composition differences and that did not include weight – a variable which needs either measurement in light indoor clothing or which is subject to documented self-reporting limitations. Although we cannot estimate the accuracy of EPIPOST in women, its generalization to females should not be straightforward as it is known that adult males have greater total lean mass and a lower fat mass than females and that these whole-body differences are complemented by major differences in tissue distribution, such as greater arm muscle mass in adult males [29].

In developing the best fitting model, we found that relaxed upper arm circumference predicted better the presence of low BMD than any other variable. Regardless of its statistical significance, upper arm circumference has also face validity as it has been suggested as a proxy of BMI and a good indicator of lean body mass depletion [36]. It has also the advantage of being easy to measure in clinical practice, requiring one piece of very portable equipment, no need for calibration and little effort from the patient, making it particularly suitable for the hasty routine of a clinical practice setting. The incorporation of variables that accounted not only for the discrimination performance and goodness-of-fit but also face validity, with known clinical value and easily obtained in a real scenario, was a major concern in our model development.

Comparing the performance of EPIPOST, OST and MORES we found that they had similar discriminatory performance as assessed by AUC but it is important to notice that EPIPOST performed slightly better and showed a wider range of likelihood ratios for different cut-offs. The likelihood ratios indicate by how much a given diagnostic test result will raise or lower the pretest probability of the target disorder. Likelihood ratios>1.0 increase the probability that the target disorder is present and likelihood ratios<1.0 decrease the probability of the target disorder. A rule to interpret likelihood has been proposed with values of >10 or <0.1 being considered to provide strong evidence to rule

in or out, respectively, the presence of the specific condition [37]. OST and EPIPOST revealed LR<0.1 at cut-offs 2 and -2, respectively, but only EPIPOST had a cut-off (EPIPOST>1) associated with a LR>10 (18.4) meaning that an individual scoring EPIPOST>1 most certainly has low BMD.

This study has the advantage of being population-based which makes generalization of the results more valid. Another advantage relates to the sample age range: men less than 65 years old represent a population with less clearly defined criteria for DXA-scan testing, in which rapid tools directing the selection for further examination are particularly useful. Many clinical societies and governmental health authorities defend global testing for men aged more than 70 years with the remaining staying in a grey zone [38]. Femoral neck was selected as reference site in conformity with the recommendation of the International Osteoporosis Foundation and also because it is has high predictive value for hip fracture risk and there is wider experience with this site [9].

The interpretation of our results must take into account several limitations. The study was conducted in a relatively small number of men, which limits statistical power: our sample of approximately 150 participants allowed for the estimation of a prevalence of 50%, with 8% precision at a 0.05 confidence level with 80% power. Because the sample was randomly selected with no clinical exclusion criteria it is possible that some osteoporotic patients have secondary causes. However, the prevalence of such causes in our cohort is likely to be very small with limited expected impact in the results. For example, none of the individuals in our sample reported to have rheumatoid arthritis. Another limitation is that we did not perform validation with an external population. However, we assessed internal validity by leave-one-out cross-validation which showed acceptable concordance.

Direct comparisons of our results with those of previous studies validating OST and MORES are difficult to make due to many methodological differences across studies in terms of sample characteristics (population source, age, race), bone mass assessment (DXA equipment and quality assessment), low bone mass diagnosis (reference population for T-score calculation, anatomical site selected). Nevertheless, a previous study assessing the performance of MORES in diagnosis osteoporosis at the hip reported an AUC of 0.842 and a sensitivity 95% of and a specificity of 61% with scores equal or higher than 6. This was a population based study including men older than 50

years with a mean age 64.2 years [23]. The performance of MORES at the lumbar spine in white men was lower with an AUC of 0.653 and a sensitivity of 51% and a specificity of 67% with a MORES≥6 [39]. Another study validating MORES in a clinical setting showed an AUC of 0.82 with a sensitivity of 80% and a specificity of 70%, at the same score, in diagnosing men with hip osteoporosis [27]. Only the latter was performed taking as reference for T-score calculation young white women from the NHANES population, the recommended gold standard which we used in our study.

The validation of OST has been performed more times than MORES. Zimering et al. [25] tested OST in a clinical sample and reported an AUC of 0.81 and a sensitivity of 85% and specificity of 51% at a cut-off of 4 in diagnosing osteoporosis at the femoral neck in white men. A population based study from Lynn et al. [22] testing OST to diagnose osteoporosis in white men at lumbar spine, femoral neck or total hip reported an AUC of 0.714 and a sensitivity of 88% and a specificity of 36% at scores equal or lower than 2. Another population based study from Portugal [40] calculated an AUC of 0.632 with a sensitivity of 77% and a specificity of 50% for OST<3 in diagnosing osteoporosis at either the femoral neck or the lumbar spine. Richards et al. [21] proposed an OST index of 6 or lower predicting osteoporosis at lumbar spine, femoral neck or total hip in white men from a clinical population with a sensitivity of 86% and a specificity of 32%. The reported AUC was 0.72. The different cut-offs proposed for OST in different settings may be regarded as confusing and highlight the need for the calibration of the scores for each population prior to their generalized use.

In this study we developed and validated a simple tool to select men for DXA testing using only relaxed upper arm circumference, which is more suitable than body weight for clinical examination, particularly following routine blood pressure determination. Our tool was similar to previously developed scores in terms of overall accuracy but showed a wider range of discriminatory ability.

References

1. Jones, G., T. Nguyen, and P. Sambrook, Symptomatic fracture incidence in elderly men and women: the Dubbo osteoporosis epidemiology study (DOES). Osteoporos Int, 1994. **4**(277-282).

2. Frost, S., Excess mortality attributable to hip-fracture: A relative survival analysis. Bone, 2013. **16**(56): p. 23-29.

3. Bliuc, D., Mortality risk associated with low-trauma osteoporotic fracture and subsequent fracture in men and women JAMA, 2009. **4**(301 (5)): p. 513-521.

4. Center, J., et al., Mortality after all major types of osteoporotic fracture in men and women: an observational study. Lancet, 1999. **13**(353 (9156)): p. 878-82.

5. Forse´n, L., A. Søgaard, and M. HE, Survival after Hip Fracture: Short- and Long-Term Excess Mortality According to Age and Gender. Osteoporos Int 1999. **10** p. 73–78.

6. Gary, M., Undertreatment of osteoporosis in men with hip fracture. Arch Intern Med, 2002. **28**(162(19)): p. 2217-22.

7. Gielen, E., et al., Osteoporosis in men. Best Practice & Research Clinical Endocrinology & Metabolism, 2011. **25**: p. 321–335.

8. Ammann, P. and R. Rizzoli, Bone strength and its determinants. Osteoporos Int, 2003. **14 Suppl 3**: p. S13-8.

9. Strom, O., et al., Osteoporosis: burden, health care provision and opportunities in the EU: a report prepared in collaboration with the International Osteoporosis Foundation (IOF) and the European Federation of Pharmaceutical Industry Associations (EFPIA). Arch Osteoporos, 2011. **6**(1-2): p. 59-155.

10. Sawka, A.M., et al., What is the Number of Older Canadians Needed to Screen by Measurement of Bone Density to Detect an Undiagnosed Case of Osteoporosis? A Population-Based Study from CaMos. Journal of Clinical Densitometry, 2006. **9**(4): p. 413-418.

11. Koh, L., et al., A simple tool to identify Asian women at increased risk of osteoporosis. Osteoporos Int, 2001. **12**: p. 699–705.

12. Cadarette, S., S. Jaglal, and T. Murray, Validation of the simple calculated osteoporosis risk estimation (SCORE) for patient selection for bone densitometry. Osteoporos Int 1999. **10**: p. 85–90.

13. Sedrine, W., et al., Development and assessment of the Osteoporosis Index of Risk (OSIRIS) to facilitate selection of women for bone densitometry. Gynecol Endocrinol, 2001. **16**: p. 245–50.

14. Lydick, E., et al., Development and validation of a simple questionnaire to facilitate identification of women likely to have low bone density. Am J Manag Care, 1998. **4**: p. 37-48.

15. Rubin, K., Comparison of different screening tools (FRAX®, OST, ORAI, OSIRIS, SCORE and age alone) to identify women with increased risk of fracture. A population-based prospective study. Bone, 2013. **10**(56(1)): p. 16-22.

16. Wildner, M., Superiority of age and weight as variables in predicting osteoporosis in postmenopausal white women. Osteoporos Int 2003. **14**: p. 950–956.

17. Michaelsson, K., Screening for osteopenia and osteoporosis: selection by body composition. Maturitas, 1996. **25**: p. 77- 82.

18. Orozco, P., Asscociations between body morphology and bone mineral density in premenopausal women. European Journal of Epidemiology 1997. **13**: p. 919–924.

19. Rud, B., The Osteoporis Self-Assessment Tool versus alternative tests for selecting postmenopausal women for bone moneral density assessment: a comparative systematic review of accuracy. Osteoporos Int 2009. **20**: p. 599–607.

20. Adler, R., M. Tran, and V. Petkov, Performance of the Osteoporosis Self-assessment Screening Tool for osteoporosis in American men. Mayo Clin Proc, 2003. **78**: p. 723–7.

21. Richards, J., et al., Validation of the Osteoporosis Self-Assessment Tool in US Male Veterans. J Clin Densitom, 2013 **12**: p. 1-6.

22. Lynn, H., et al., An evaluation of osteoporosis screening tools for the osteoporotic fractures in men (MrOS) study. Osteoporos Int 2008. **19**: p. 1087–1092.

23. Shepherd, A., et al., Development and Internal Validation of the Male Osteoporosis Risk Estimation Score. Ann Fam Med 2007. **5**: p. 540-546.

24. Lynn, H., et al., An osteoporosis screening tool for Chinese men. Osteoporos Int 2005. **16**: p. 829–834.

25. Zimering, M., et al., Validation of a Novel Risk Estimation Tool for Predicting Low Bone Density in Caucasian and African American Men Veterans. Journal of Clinical Densitometry, 2007. **10**: p. 289-297.

26. Pongchaiyakul, C. and E. Wanothayaroj, Performance of the Khon Kaen Osteoporosis Study (KKOS) Score for Identifying Osteoporosis in Men. J Med Assoc Thai, 2007. **90** p. 1518-23.

27. Cass, A. and A. Shepherd, Validation of the Male Osteoporosis Risk Estimation Score (MORES) in a Primary Care Setting. JABFM J, 2013. **26**: p. 436-444.

28. Riggs, B., et al., A Population-Based Assessment of Rates of Bone Loss at Multiple Skeletal Sites: Evidence for Substantial Trabecular Bone Loss in Young Adult Women and Men. J Bone Miner Res, 2008 **23**: p. 205–214.

29. Wells, J., Sexual dimorphism of body composition. Best Practice & Research Clinical Endocrinology & Metabolism, 2007. **21**: p. 415–430.

30. Ramos, E., C. Lopes, and H. Barros, Investigating the effect of nonparticipation using a population-based case-control study on myocardial infarction. Ann Epidemiol, 2004. **14**(6): p. 437-41.

31. Gibson, R.S., (ed), Principals of Nutritional Assessment. 2nd edn Oxford University Press: New York, 2005.

32. WHO, Assessment of fracture risk and its application to screening for postmenopausal osteoporosis. Report of a WHO Study Group. WHO Technical Report Series 1994. **843**(World Health Organization, Geneva).

33. Kanis, J., et al., A reference standard for the description of osteoporosis. Bone, 2008. **42**: p. 467–75.

34. Kelly, T., K. Wilson, and S. Heymsfield, Dual Energy X-Ray Absorptiometry Body Composition Reference Values from NHANES. PLoS ONE, 2009(9): p. e7038.

35. Kanis, J., et al., The diagnosis of osteoporosis. J Bone Miner Res 1994. **9**: p. 1137–41.

36. James, W., et al., The value of arm circumference measurements in assessing chronic energy deficiency in Third World adults Eur J Clin Nutr 1994. **48**(12): p. 883-894.

37. Jaeschke, R., G. Guyatt, and J. Lijmer, Diagnostic Tests, ed. AmericanMedicalAssociation. 2002.

38. Tavares, V., et al., Recommendations for the diagnosis and management of osteoporosis. Acta Reumatol Port. , 2007 **32**(1): p. 49-59.

39. Shepherd, A.J., A.R. Cass, and L. Ray, Determining risk of vertebral osteoporosis in men: validation of the male osteoporosis risk estimation score. J Am Board Fam Med, 2010. **23**(2): p. 186-94.

40. Machado, P., M. Coutinho, and J.A. da Silva, Selecting men for bone densitometry: performance of osteoporosis risk assessment tools in Portuguese men. Osteoporos Int, 2010. **21**(6): p. 977-83.

Table 1. Study sample characteristics: overall and by bone mineral density category

Baseline Characteristics	**Men**			
(mean, *sd*)	(*n* = 147)	**Low BMD** (*n* = 49)	**Normal BMD** (*n* = 98)	***p*-value**
Age (years)	53.6 (7.03)	54.8 (6.96)	52.9 (7.02)	0.123
Weight (Kg)	77.8 (12.97)	70.9 (10.81)	81.3 (12.61)	<0.001
Height (cm)	171.4 (7.16)	170.3 (6.14)	172.0 (7.59)	0.180
BMI (kg/m^2)	26.4 (3.66)	24.4 (3.04)	27.4 (3.53)	<0.001
Circumferences				
Relaxed upper arm (cm)	31.7 (3.01)	29.8 (2.81)	32.6 (2.67)	<0.001
Flexed upper arm (cm)	32.6 (3.00)	30.8 (2.71)	33.5 (2.73)	<0.001
Waist (cm)	96.3 (10.41)	90.7 (9.55)	99.2 (9.67)	<0.001
Abdomen (cm)	97.8 (10.23)	92.3 (9.24)	100.5 (9.62)	<0.001
Hip (cm)	99.5 (6.74)	96.2 (5.64)	101.2 (6.66)	<0.001
Thigh (cm)	52.2 (3.90)	50.0 (3.56)	53.3 (3.60)	<0.001
DXA				
Lumbar Spine BMD (g/cm^2)	− 0.969 (0.140)	0.889 (0.127)	1.009 (0.130)	<0.001
Lumbar Spine T-score	1.071 (1.246)	− 1.785 (1.139)	− 0.720 (1.147)	<0.001
Femoral Neck BMD (g/cm^2)	0.794 (0.1243	0.662 (0.050)	0.859 (0.094)	<0.001
Femoral Neck T score	− 0.537 (1.036)	− 1.636 (0.417)	0.012 (0.784)	<0.001

BMI, body mass index; DXA, Dual-energy X-ray absorptiometry.

Table 2. Crude and Adjusted Analysis of Age and different anthropometric variables and low bone mineral density

	Crude Analysis		Model 1*	
Variables	**OR**	**95% CI**	**OR**	**95% CI**
Age	1.04	0.99-1.09	1.03	0.98-1.09
Weight	0.92	0.89-0.96	0.98	0.93-1.04
Height	0.97	0.92-1.02	1.01	0.95-1.06
BMI	0.73	0.64-0.84	0.90	0.72-1-13
Relaxed UAC	0.68	0.57-0.79		
Waist	0.91	0.87-0.95	0.97	0.91-1.04
Abdomen	0.91	0.87-0.95	0.98	0.91-1.05
Hip	0.87	0.82-0.93	0.98	0.89-1.08
Thigh	0.76	0.67-0.86	0.91	0.76-1.08

OR, Odds ratio; CI, Confidence interval; BMI, Body mass index; UAC, upper arm circumference.

* Model adjusted for relaxed UAC

Table 3. Global Fit and Accuracy of the Different Tools before and after calibration

Screening Tools	Hosmer- Lemeshow GOF test (*p* value)	AUC
OST	<0.001	0.73
OST $_{\text{Calibrated}}$	0.244	0.73
MORES	0.002	0.75
MORES $_{\text{Calibrated}}$	0.663	0.75
EPIPOST	0.631	0.76

OST, Osteoporosis Screening Tool; MORES, Male Osteoporosis Risk Estimation Score; EPIPOST, EpiPorto Osteoporosis Screening Tool; GOF, Goodness of fit test; AUC, area under the curve.

Table 4. Likelihood Ratios for Different cut-Offs of Calibrated OST, Calibrated MORES and EPIPOST

	Range	Prevalence	Low BMD		LR	AUC
		n (%)	Yes (n (%))	No (n (%))		
OST $_{Calibrated}$*	**≤ -2**	8 (5.4)	0 (0.0)	8 (8.2)	0.0	0.73
	]-2, -1]	72 (49.0)	15 (30.6)	57 (58.2)	0.5	
	]-1, 0]	43 (29.3)	21 (42.9)	22 (22.4)	1.9	
	>0	24 (16.3)	13 (26.5)	11 (11.2)	2.4	
Total		**147 (100)**	**49 (100)**	**98 (100)**		
MORES $_{Calibrated}$	**≤ -2**	33 (22.4)	3 (6.1)	30 (30.6)	0.2	0.75
	]-2, -1]	27 (18.4)	3 (6.1)	24 (24.5)	0.3	
	]-1, 0]	46 (31.3)	19 (38.8)	27 (27.6)	1.4	
	>0	41 (27.9)	24 (49.0)	17 (17.3)	2.8	
Total		**147 (100)**	**49 (100)**	**98 (100)**		
EPIPOST	**≤ -2**	19 (13.0)	1 (2.0)	18 (18.6)	0.1	0.76
	]-2, -1]	48 (32.9)	9 (18.4)	39 (40.2)	0.5	
	]-1, 0]	47 (32.2)	18 (36.7)	29 (29.9)	1.2	
	]0, 1]	22 (15.1)	12 (24.5)	10 (10.3)	2.4	
	>1	10 (6.8)	9 (18.4)	1 (1.0)	18.4	
Total		**146 (100)**	**49 (100)**	**97 (100)**		

BMD, Bone Mineral Denstiry; LR, Likelihood ratio; AUC, area under the receiver operating characteristic curve; OST, Osteoporosis Screening Tool; MORES, Male Osteoporosis Risk Estimation Score; EPIPOST, EpiPorto Osteoporosis Screening Tool

* To simplify results interpretation, OST results are presented inverted.

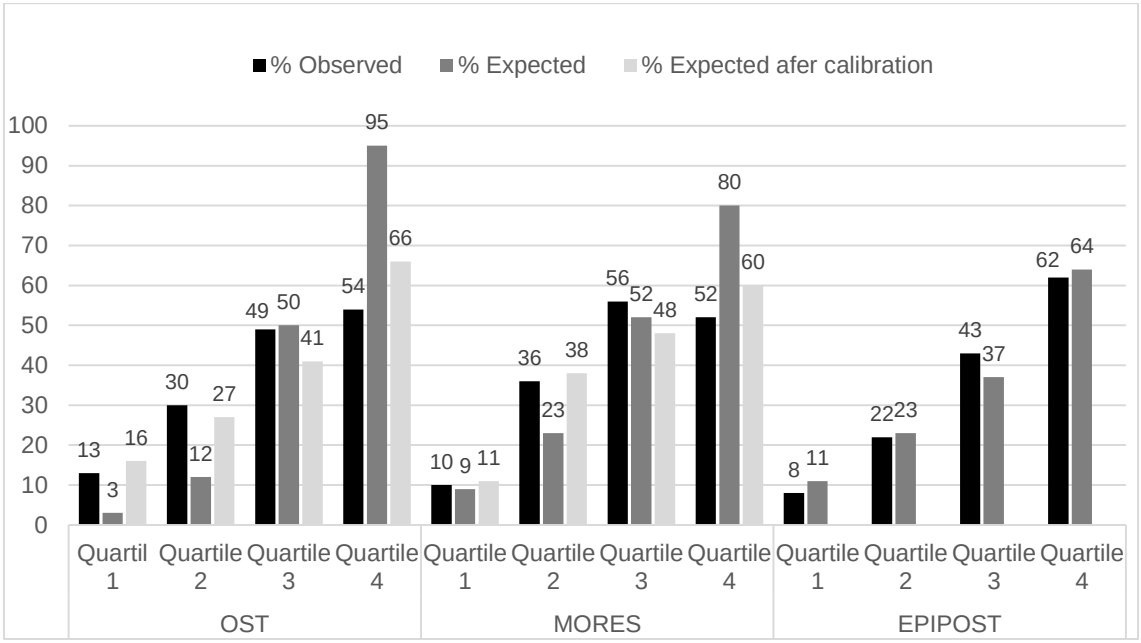

Figure 1. Calibration of MORES ans OST: prevalence of low bone mineral density before and after calibration through quartiles of the different scores. To simplify results interpretation, OST results are presented inverted.

Appendix 1

$OST = \beta_0 + 0.2$ (body weight in Kg−age in years)

$OST = 3.03 + 0.2$ (body weight in Kg−age in years)

$OST_{calibrated} = \beta_0 + \beta_1$ (body weight in Kg−age in years)

$OST_{calibrated} = 0.63 + 0.33$ (body weight in Kg−age in years)

$MORES = \beta_0 + \beta_1$ (Age 56-65 years) + β_2 (Weight≤70 Kg) + β_3 (80Kg>Weight>70Kg) + β_4*COPD

MORES = -3.02 + 1.29 (Age 56-65 years) + 3.07 (Weight≤70 Kg) + 1.86 (80Kg>Weight>70Kg) + 1.32*COPD

$MORES_{calibrated} = \beta_0 + \beta_1$ (Age 56-65 years) + β_2 (Weight≤70 Kg) + β_3 (80Kg>Weight>70Kg) + β_4*COPD

$MORES_{calibrated}$ = -2.47 + 0.70 (Age 56-65 years) + 2.01 (Weight≤70 Kg) + 2.13 (80Kg>Weight>70Kg) + 1.32*COPD

CONCLUSION

Our systematic review identified 22 studies validating 9 tools aimed to identify men with low BMD benefiting from DXA testing. However, of these, only 3 (OST, OSTA and MORES) were validated in more than in one sample. Despite high methodological heterogeneity across studies, tools screening for low BMD in men seem to have acceptable accuracy but no tool performed consistently better than other. OST and MORES, for their simplicity, accuracy and replication, seem to be more adequate for routine clinical practice use.

Calibration of OST and MORES improved the fit of both models to our population data while maintaining their discriminatory ability to identify men with low bone mineral density. The newly developed tool specifically for men, EPIPOST, proved to have slightly better discriminatory ability than OST and MORES while being simpler to execute. The likelihood ratios analysis revealed that individuals scoring OST≤2, MORES>-2 and EPIPOST>-2 should undergo DXA testing.

REFERENCES

1. U.S. Department of Health and Human Services. Bone Health and Osteoporosis: A Report of the Surgeon General. Rockville, MD: U.S. Department of Health and Human Services; 2004.

2. Nelson, H., et al., Screening for Osteoporosis: Systematic Review to Update the 2002 U.S. Preventive Services Task Force Recommendation. Rockville (MD): Agency for Healthcare Research and Quality (US), 2010 Jul.

3. Strom, O., et al., Osteoporosis: burden, health care provision and opportunities in the EU: a report prepared in collaboration with the International Osteoporosis Foundation (IOF) and the European Federation of Pharmaceutical Industry Associations (EFPIA). Arch Osteoporos, 2011. **6**(1-2): p. 59-155.

4. Bleibler, F., et al., The health burden and costs of incident fractures attributable to osteoporosis from 2010 to 2050 in Germany—a demographic simulation model. Osteoporos Int 2013 **24**: p. 835–847.

5. Schuit, S.C., et al., Fracture incidence and association with bone mineral density in elderly men and women: the Rotterdam Study. Bone, 2004. **34**(1): p. 195-202.

6. Giangregorio, L., et al., Fragility fractures and the osteoporosis care gap: an international phenomenon. Semin Arthritis Rheum, 2006. **35**(5): p. 293-305.

7. Johnell, O., et al., Mortality after osteoporotic fractures. Osteoporos Int, 2004. **15**(1): p. 38-42.

8. Forsén, L., et al., Survival after Hip Fracture: Short- and Long-Term Excess Mortality According to Age and Gender. Osteoporos Int 1999 **10**: p. 73–78.

9. Hopkins, R.B., et al., Estimation of the lifetime risk of hip fracture for women and men in Canada. Osteoporos Int, 2012. **23**(3): p. 921-7.

10. Kannegaard, P.N., et al., Excess mortality in men compared with women following a hip fracture. National analysis of comedications, comorbidity and survival. Age Ageing, 2010. **39**(2): p. 203-9.

11. Papaioannou, A., et al., The osteoporosis care gap in men with fragility fractures: the Canadian Multicentre Osteoporosis Study. Osteoporos Int, 2008. **19**(4): p. 581-7.

12. Ebeling, P.R., Osteoporosis in men. Curr Opin Rheumatol 2013. **25**: p. 542–552.

13. Melton, L., et al., Fractures Attributable to Osteoporosis: Report from the National Osteoporosis Foundation. Journal of Bone and Mineral Research, 1997. **12**: p. 16-23.

14. Warrinera, A., N. Patkarb, and J. Curtisb, Which fractures are most attributable to osteoporosis? Journal of Clinical Epidemiology, 2011. **64** (1): p. 46–53.

15. Ammann, P. and R. Rizzoli, Bone strenght and its determinants. Osteoporosis International, 2003. **14**(Suppl 3): p. 13-18.

16. Kanis, J., et al., The diagnosis of osteoporosis. J Bone Miner Res 1994. **9**: p. 1137–41.

17. Marshall, D., O. Johnell, and H. Wedel, Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures Br Med J, 1996. **312**: p. 1254–9.

18. Kanis, J., et al., A reference standard for the description of osteoporosis. Bone, 2008. **42**: p. 467–75.

19. Machado, P., M. Coutinho, and J.A. da Silva, Selecting men for bone densitometry: performance of osteoporosis risk assessment tools in Portuguese men. Osteoporos Int, 2010. **21**(6): p. 977-83.

20. National Osteoporosis Foundation. Clinician's Guide to Prevention and Treatment of Osteoporosis. Washington, DC: National Osteoporosis Foundation, 2010.

21. Tavares, V., et al., Recommendations for the diagnosis and management of osteoporosis. Acta Reumatol Port. , 2007 **32**(1): p. 49-59.

22. Schousboe, J.T., et al., Cost-effectiveness of bone densitometry among Caucasian women and men without a prior fracture according to age and body weight. Osteoporos Int, 2013. **24**(1): p. 163-77.

23. McGinn, T.G., et al., Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. Jama, 2000. **284**(1): p. 79-84.

24. Shepherd, A., et al., Development and Internal Validation of the Male Osteoporosis Risk Estimation Score. Ann Fam Med 2007. **5**: p. 540-546.

25. Liu, H., N.M. Paige, and C. Goldzweig, Screening for Osteoporosis in Men: A Systematic Review for an American College of Physicians Guideline. Annals of Internal Medicine, 2008. **148**(9): p. 685-701.

26. Ng, S.P., et al., Validity of self-reported height and weight and derived body mass index in middle-aged and elderly individuals in Australia. Aust N Z J Public Health, 2011. **35**(6): p. 557-63.