

# Assessing risk of osteoporotic fractures in primary care: development and validation of the FRA-HS algorithm

**Lapi Francesco<sup>1</sup>, Bianchini Elisa<sup>1</sup>, Michieli Raffaella<sup>2</sup>, Pasqua Alessandro<sup>1</sup>, Cricelli Iacopo<sup>1</sup>, Mazzaglia Giampiero<sup>1</sup>, Frediani Bruno<sup>3</sup>, Prieto-Alhambra Daniel<sup>4,5</sup>, Brandi Maria Luisa<sup>6</sup>, Cricelli Claudio<sup>2</sup>**

1. *Health Search, Italian College of General Practitioners and Primary Care, Florence (Italy)*
2. *Italian College of General Practitioners and Primary Care, Florence (Italy)*
3. *Department of Rheumatology, Policlinico Le Scotte, University of Siena, Siena (Italy)*
4. *Oxford NIHR Musculoskeletal Biomedical Research Unit, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford (United Kingdom)*
5. *GREMPAL Research Group, Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Universitat Autònoma de Barcelona, Barcelona (Spain)*
6. *Metabolic Bone Diseases Unit, Department of Surgery and Translational Medicine, University of Florence, Florence (Italy)*

**Running title:** FRA-HS to predict osteoporotic fractures

## **Corresponding author:**

Francesco Lapi, PharmD, PhD

*Health Search, Italian College of General Practitioners and Primary Care*

Via Sestese 61

50141 Florence, Italy

Phone number: +39 055 494900 or +39 3402648088

Fax: +39 055 494900

email: lapi.francesco@simg.it

All authors, meeting the criteria for authorship as established by the International Committee of Medical Journal Editors, have read and approved the manuscript. The manuscript is not under consideration for publication elsewhere and all named authors have agreed to its submission.

**Keywords:** osteoporotic fractures, FRA-HS, score, primary care

## **Acknowledgements**

This study has been funded by the Italian College of General Practitioners and Primary Care.

## Abstract

We aimed to develop and validate the FRActure Health Search (FRA-HS) score for prediction of risk of osteoporotic fractures in primary care in Italy. We selected a cohort of patients aged 40 years between 1999 and 2002. They were followed until the occurrence of osteoporotic fracture, death, end of data registration, or end of data availability (December 31, 2012). Age, sex, history osteoporotic fractures, secondary osteoporosis, long-term use of corticosteroids, rheumatoid arthritis, body mass index, smoking, and alcohol abuse/alcohol-related diseases, and the interaction terms sex\*use of corticosteroids and age\*secondary osteoporosis were entered in a competing-risk regression (Fine & Gray method) to predict the risk of hip/femur or overall major osteoporotic fractures. The coefficients were combined to obtain the FRA-HS for individual patients. Explained variance, discrimination and calibration measures were computed to evaluate the models accuracy. The final model was tested using an independent data source. The FRA-HS explained 47.36% and 20.6% of the variation for occurrence of hip/femur and overall major osteoporotic fractures, respectively. Area Under Curve was 0.77 and 0.73, respectively. Predicted/observed ratios revealed a margin of error lower than 30% in the 80% of the population. After stratifying by sex, prediction models for hip/femur fractures confirmed acceptable accuracy in both sexes, while poor explained variance (<20%) was observed for overall major fractures. These findings indicate that FRA-HS might be implemented in primary care for risk prediction of hip/femur fractures. General practitioners could be therefore supported by this tool in clinical decision making.

## Introduction

Osteoporotic fractures are a major and growing concern for patients, physicians and public health systems.[1–7] Given that several lifestyle and treatment interventions might reduce the risk of osteoporotic fractures, several prediction algorithms have been developed on the bases of demographic and clinical risk factors.[8] Among the available tools, the most used worldwide is the FRAX® score, a system incorporated in several guidelines in western countries, to predict the 10-year probability of osteoporotic fractures. The FRAX® supports clinicians in stratifying the risk and deciding tailored therapy for each patient irrespective of Body Mass Density (BMD). Indeed, the patient's risk assessment limited to BMD value was demonstrated as poorly sensible and expensive.[9–11]

As other prediction models, it is pivotal to evaluate FRAX® performance in populations representative of the clinical setting in which this tool will be applied. Indeed, the distribution of demographics and clinical risk factors may vary across different countries or settings along with the incidence rate of osteoporotic fractures and mortality data, which are variables deeply affected by over-time changes of healthcare and socioeconomic factors.[1, 12–14]

In Italy, the first implementation of FRAX® was performed, in 2008, using hip fractures rates registered in four cities and one region. These data stemmed from claims databases ranging 1994-2008 for what concerns fractures rates, while mortality rates were based on 1999 estimates.[12] Subsequently, Piscitelli and coworkers[15] updated the Italian version of FRAX® with hospital records of osteoporotic fractures and morality data referring to 2008. This analysis was carried out using the national administrative database managed by the Italian Ministry of Health.

Nevertheless, these prior studies were conducted using claims databases, whereby osteoporotic fractures were those registered as hospital admissions. As such, while hip fractures were accurately registered, the other major osteoporotic fractures (i.e., vertebral, humeral or wrist/forearm fractures), whose minority requires hospitalization, were likely featured by less accuracy. Furthermore, some clinical risk factors adopted to calculate the Italian FRAX®, such as BMI and alcohol consumption, were derived from prior investigations[16] given their absence in administrative data.[15]

In this context, primary care data, which stem from the Italian general practitioners (GPs), have the advantage of including large representative populations with several years of follow-up, lifestyle and clinical variables, which are dynamically updated over time and are not limited to hospitalization records.

We therefore developed and validated the FRActure Health Search (FRA-HS) score, a FRAX®-based model, for the assessment of risk of osteoporotic fractures in primary care in Italy.

## **Methods**

### **Data source**

We developed the FRA-HS using the Health Search – IMS Health Longitudinal Patients Database (HSD), an Italian general practice database that includes patients' records of a group of over 1000 GPs homogeneously distributed across Italy. GPs voluntarily agreed to collect patients' information and to attend training courses for data entry. Patients' demographic details are linked with an encrypted code to clinical records (diagnoses, referrals, tests prescriptions and results), drug prescriptions (drug name, date of filled prescription, and number of days' supply), lifestyle-related records (i.e., Body Mass Index (BMI), smoking and alcohol consumption), hospital admissions, and date of death. Data within the HSD is coded using internationally recognized codes, such as the 9th version of the International Classification of Disease, Clinical Modification (ICD-9-CM) for medical diagnoses and the Anatomical Therapeutic Chemical (ATC) classification system for drugs.

To be considered for participation in epidemiological studies, GPs are required to meet standard quality criteria pertaining to the levels of coding, prevalence of well-known diseases, mortality rates, and years of recording.[17] When this study initiated, 700 GPs homogeneously distributed across all areas of Italy, and covering a patient population of 1,163,855 individuals, met the standard quality criteria and were included in the study. The validity of the HSD data for conducting epidemiological research has been confirmed elsewhere.[18–20]

The “Mille in Rete” database was adopted as independent (external) validation cohort. This is a local (Veneto region, in the North of Italy) database formed by 62 GPs who do not belong to HSD network, and it is adopted by the local health authority to perform public health evaluations. “Mille in Rete” data are coded using the same internationally recognized thesaurus of HSD, and its reliability has been already demonstrated elsewhere.[21]

### **Study population**

We formed a cohort of patients aged 40 years and over during the period between January 1, 1999 and December 31, 2002. To be considered eligible, patients were required to have at least 1-year medical history in the database. Given that HSD is a dynamic database, patients'

data are longitudinally recorded and available up to December 31, 2012, when this study was conducted. By doing so, patients could be potentially followed up for 10 years. The date of the first GP's visit within the eligibility period was the study index date. Each patient of the selected cohort was therefore followed until the occurrence of these events, whichever came first: osteoporotic fracture (event date), death from any cause, end of registration with GP, or end of the study period (December 31, 2012). We adopted the same selection criteria to form the validation cohort using "Mille in Rete" database. However, the baseline eligibility period ranged from January 1, 2003 to December 1, 2005 (the end of data is December 31, 2015 for "Mille in Rete") so that this population was also temporally distinct from the developmental population as indicated by external validation methods for prediction modeling.[22–24]

### **Outcome definition**

We identified all diagnoses coded via ICD9CM which were consistent with osteoporotic fractures occurred during follow-up. Namely, hip/femur (**ICD9CM code:** 820\*, 821.0, 821.2), vertebral (**ICD9CM code:** 805\*), humeral (**ICD9CM code:** 812\*), wrist/forearm (**ICD9CM code:** 813\*) fractures were defined as the study outcome. These codes have been selected on the bases of previous consensus studies in which the relationship between osteoporosis and fracture sites coded via ICD9CM have been assessed. Namely, they excluded codes referring to multiple, open and facial/skull fractures (i.e. likely traumatic fractures), which are generally considered as least likely related to osteoporosis.[25, 26] In addition, we did not include "pathologic" fractures (some ICD9CM sub-codes of 733\*) in the outcome because of their potential relationship with malignant neoplasia in place of osteoporosis.[27] In line with this approach, these codes have been adopted in previous database studies on osteoporotic fractures.[15, 18, 28]

### **Risk factors**

We considered all clinical risk factors, along with age (centered on the population mean age) and sex, known to affect the development of osteoporotic fracture[9, 29] according to FRAX® score. They comprised history of osteoporotic fractures (including all the aforementioned ICD9CM codes along with pelvis (**ICD9CM code:** 808\*) and tibia/fibula (**ICD9CM code:** 823\*, for women only) fractures, other causes of secondary osteoporosis (type I diabetes (**ICD9CM code:** 250.\*1, 250.\*3; osteogenesis imperfecta (**ICD9CM code:** 756.51), hyperthyroidism (**ICD9CM code:** 242.0, 242.1, 242.8, 242.9), hypogonadism (**ICD9CM code:** 256.39, 257.2) or premature menopause (**ICD9CM code:** 627.9), chronic malnutrition or malabsorption

(**ICD9CM code:** 555, 556, 579) and chronic liver disease (**ICD9CM code:** 570, 571)), chronic use of corticosteroids (**ATC code** H02\* and at least 180 Defined Daily Dose (DDD) within one year before the index date[30]), rheumatoid arthritis (**ICD9CM code:** 714\* and 720.0), Body Mass Index (BMI), current smoking, and alcohol abuse (i.e., >40 g and >20 g daily for men and women, respectively) or alcohol-related diseases (**ICD9CM code:** 291, 303, 305.0, 357.5, 535.3, 790.3). Sex and use of corticosteroids, along with age and secondary osteoporosis entered the model as interaction terms [9, 29]. Each risk factor was defined before or on the index date.

## Data analysis

We reported descriptive statistics for continuous (mean (SD)) and categorical values (% and related 95% Confidence Intervals (CI)). We calculated the incidence rate of hip/femur and overall major fractures dividing the number of events by the person-times cumulated during follow-up.

We entered age and sex, and all the aforementioned clinical risk factors, in multivariate competing-risk regression models, where the outcome was modelled accounting for a competing risk with mortality.[31, 32] In the primary analysis, we used models with missing categories (i.e. missing values were imputed with a fictitious category “999”) for BMI and smoking. We therefore estimated **Model 1** and **Model 2** with hip/femur and overall major fractures as outcome, respectively. We opted for these two primary models because of two main reasons. First, exclusion of patients with missing values sensibly reduced the statistical power (30232 vs. 439 cases of fracture). Second, smoking and BMI had several missing values (>90%) to consider the primary model as that with ‘imputed’ missing data. However, given the size of patients with missing values, we also reported the results obtained through models (**Model 3**, **Model 4**) which excluded patients with missing values.

Each risk factor was assigned a regression b coefficient estimated in the models, and a patient-specific score was computed through the linear combination of individual coefficients, excluding those estimated for missing categories. The FRA-HS was therefore categorized in deciles. We evaluated the accuracy of the FRA-HS score by calculating the explained variance (pseudo R<sup>2</sup>) as a performance measure, the D-statistic and Area Under the Curve (AUC) as discrimination measures, and the ratio between predicted and observed fractures as a calibration measure (a ratio of 1 indicates perfect calibration).[23] As confirmatory analysis, we evaluated the presence of ‘moderate’ calibration using flexible calibration curve. Namely, we used a spline-based tool with confidence intervals to plot predicted vs. observed risk of

osteoporotic fractures.[33, 34] To ensure the highest statistical power, this analysis was limited to overall major fractures.

We stratified the results by sex. For these analyses, the models did not include the interaction term with sex.

The accuracy of **Model 1** and **Model 2** was verified in an independent (external) primary care database (“Mille in Rete” of Veneto, a region of Northern Italy) by applying the FRA-HS to this cohort and calculating the related **pseudo R<sup>2</sup>**, AUC, and the predicted/observed ratios.

We calculated the 10-year predicted risk of fractures by taking the regression coefficients from the primary models and using them as weights for FRA-HS. Namely, we combined these weights with the baseline survivor function for the diagnosis of osteoporotic fractures obtained from the competing-risk multivariable model evaluated at 10 years (**Appendix 1**). This same approach has been used previously.[35]

Given that models with missing categories may be biased by uncertainty resulting from missingness [36–40], we conducted a sensitivity analysis to evaluate the burden of missing data on the results. According to methodological literature [39, 41, 42], we replaced missing values for BMI and smoking status using Multiple Imputation (MI) procedure. In practical, we created five copies of the dataset and replaced missing values with imputed values based on suitable random sample from their predicted distributions. Prediction was obtained using the other available variables (i.e., gender, age, history of fractures, chronic use of corticosteroids, rheumatoid arthritis and secondary osteoporosis) along with the outcome variables as per Moons and coworkers.[43] We therefore estimated two further models (**Model 5**, **Model 6**) using the dataset with imputed missing values for both hip/femur and overall osteoporotic fractures. The **pseudo R<sup>2</sup>**, D-statistic, AUC, and the predicted/observed ratio were calculated for these models as well.

## Results

**Table 1** depicts the characteristics of the cohorts including (n=407,771) or excluding (n=6,893) patients with missing data for BMI and/or smoking. In details, mean age was around 60 years in both cohorts, while there were more (55.04%) and less (46.2%) females than males for the cohorts excluding or including missing data, respectively. The other variables showed similar distribution in the two cohorts, with the exception of history of osteoporotic fractures (1.27 vs. 1.65%; in males (0.94 vs. 1.21%) and females (1.54 vs. 2.17%)), whose prevalence was higher for patients with non-missing values, along with

alcohol abuse and long-term use of corticosteroids, which were proportionally higher in males (0.66 vs. 1.46%) and females (2.58 vs. 2.86) with non-missing values, respectively. Furthermore, a little more than 40% of patients were overweight, 30% were obese or with normal weight, and less than 1% were underweight; 30% were current smokers; those with rheumatoid arthritis and alcohol abuse/alcohol-related diseases were below 1%, while secondary osteoporosis and chronic use of corticosteroids ranged 2-4%.

**Table 2** displays the baseline characteristics of patients whether they experienced or not osteoporotic fractures during follow-up. Patients who experienced osteoporotic fractures were generally older, most of them were women, had a higher BMI and suffered from rheumatoid arthritis, secondary osteoporosis, alcohol related diseases or were alcohol abusers, and were more likely exposed to long-term use of corticosteroids. Only current smokers were proportionally higher among those who did not experience an osteoporotic fracture compared with fractured patients.

**Table 3** depicts the incidence rates for hip/femur and other osteoporotic fractures estimated for the selected cohort. Overall, females showed higher rates than males for all osteoporotic fractures. In specific, rates of hip/femur, humeral, and wrist/forearm fractures were two-fold or greater in women than men. The mean age on fracture date was higher than sixties for all type of fractures. It was higher for hip/femur fractures than overall major fractures (77.1 vs. 66.4 in men; 80.4 vs. 69.6 in women) and for females than males.

The FRA-HS score obtained with **Model 1** and **Model 2** explained 47.36% (95% CI: 47.11-47.84) and 20.6% (95% CI: 20.34-21.06) of the variation for hip/femur and overall major fractures occurrence, respectively. In terms of discrimination, AUC was 0.77 (95% CI: 0.76-0.78) and 0.64 (95% CI: 0.63-0.65) for **Model 1** and **Model 2**, respectively. **Model 3** and **Model 4** reported greater explained variation (52.58% (95% CI: 46.98-57.43); 21.7% (95% CI: 18.89-23.76, respectively)), but lower discrimination values for both hip/femur (AUC=0.73 (95% CI: 0.66-0.80)) and overall major osteoporotic fractures (AUC=0.65 (95% CI: 0.61-0.69)), respectively. When we stratified the results by sex, both explained variance and discrimination measures were greater for females than males for both hip/femur (**Model 1**: **pseudo R<sup>2</sup>**= 42.69 vs. 34.58%; AUC=0.76 vs. 0.66) and overall major fractures (**Model 2**: **pseudo R<sup>2</sup>**=15.9 vs. 5.25%; AUC=0.62 vs. 0.49) (**Table 4**).

Concerning calibration, for **Model 1** the margin of error was  $\leq 34\%$  (underestimation) in the 80% of the population, except for the 3<sup>rd</sup> (57%) and 4<sup>th</sup> (47%) decile. With **Model 2** the margin of error was  $\leq 30\%$  in the 90% of the population. When we stratified the analysis by



sex, women generally showed better calibration than men. Similar results were observed for **Model 3** and **Model 4** when compared with primary models (**Table 5**).

**Fig. 1** depicts the flexible curve calibration. The spline regression fitting predicted versus observed risk of overall major osteoporotic fractures, showed some over and underestimation. However, the confidence intervals overlapped the line of perfect calibration for most of the FRA-HS values as confirmed by calibration slope (1.00 (95% CI: 0.83-1.18)), calibration intercept (0.16 (95% CI: 0.03-0.29)), and discrimination value (c-statistics: 0.72 (95% CI: 0.68-0.75)).

Mathematical formulas of prediction **Model 1** and **Model 2** are reported in **Appendix 1**.

When we estimated the **Model 1** and **Model 2** using “Mille in Rete” database as validation cohort (N=82242; mean age: 57.7 (SD: 12.07); 55.24% women), explained variance, discrimination and calibration measures were generally consistent with or even improved those obtained in developmental population. Namely,  $R^2$  and AUC were higher for hip/femur fractures (48.26 (95% CI: 49.28-50.12; 0.85 (95% CI: 0.84-0.86)) than overall major osteoporotic fractures (22.97 (95% CI: 21.37-23.97); 0.70 (95% CI: 0.69-0.71)) (**Table 6**). Concerning calibration, a certain overestimation for hip/femur fractures was observed from the 5<sup>th</sup> decile and above, especially among men. In women, the margin of error was  $\leq 38\%$  in the 90% of the population for the risk of hip/femur fractures. In respect of overall major fractures, the margin of error was  $\leq 38\%$  and  $\leq 31$  among men and women, respectively (**Table 7**). As seen in HSD cohort, FRA-HS has a better predictive accuracy in women than men. Overall, there was 2.4% 10-year predicted risk of hip/femur fractures, which was 3-fold higher in females than males (3.4 vs. 1.3%).

**Fig. 2** shows the contribution to FRA-HS of individual clinical risk factors provided by **Model 1** and **Model 2** concerning prediction of 10-year risk of fractures. Namely, when we considered men and women aged 60-70 years with normal weight, the role of individual factors for hip/femur and overall major osteoporotic fractures were generally greater in women than men. Nevertheless, secondary osteoporosis had almost the same effect in men and women for hip/femur fractures, and chronic use of corticosteroids had a higher burden in women than men for any kind of fractures.

**Fig. 3** shows the results obtained with **Model 3** and **Model 4**. The role of individual factors for hip/femur and overall osteoporotic fractures was generally greater in women than men, even without any clinical risk factor. Concerning the risk of hip/femur fractures, alcohol abuse had the highest burden in men, while there was no registration in women. Secondary

osteoporosis was the most relevant factor among women. Considering all major fractures, the chronic use of corticosteroids had the greatest weight in increasing the risk of fractures among men. It also has more relevance in men than women along with secondary osteoporosis.

Details on individual coefficients estimated through all models are reported in **Supplemental Table 1**. Notably, the primary models showed an increased risk for all included variables with the exception of being obese or underweight for **Model 1**. Instead, **Model 3** and **Model 4** as well as the two models estimated in the sensitivity analysis reported some “protective” hazard ratios, so indicating a reduced risk of fractures for certain patients’ features.

Overall, the extent of each clinical risk factor increased with age for all models (**Supplemental Table 2** and **Supplemental Table 3**).

In the sensitivity analysis, when we imputed missing values, the multivariable models showed an accuracy roughly consistent with those obtained for the primary models (see **Supplemental Table 4** and **Supplemental Table 5**). Given the relevant burden of missing values, we evaluated the performance of MI technique by comparing the datasets (n=5) with imputed missing values with the full dataset. As shown in **Supplemental Figure 1** there was concordance between full and imputed datasets for both BMI and smoking categories.

## Discussion

This study reported the accuracy level of FRA-HS score for predicting osteoporotic fractures in primary care. Overall, the model had acceptable accuracy in terms of overall performance, discrimination and calibration. When we stratified the analysis by sex, FRA-HS reliably performed in terms prediction of hip/femur fractures for both men and women, while its predictive accuracy needs to be improved for overall major osteoporotic fractures. When the model was estimated using imputed missing data as a sensitivity analysis, the results were generally consistent with those yielded for the primary models.

We chose to develop the FRA-HS on the bases of FRAX® model, given that the accuracy of this score has been demonstrated by 26 population-based studies conducted in 9 countries, and it is currently incorporated in several treatment guidelines worldwide.[10, 13, 44, 45] Other 13 similar prognostic tools have been developed, but only Garvan, and QFracture score have been validated in at least 2 population-based studies using an identical definition for the algorithm. Nevertheless, Garvan and QFracture score have been evaluated in only 1 and 3 countries, respectively, and they did estimate the fracture risk without taking into account the

competing risk with death.[10, 14, 31, 32, 46, 47] Although a higher number (n=19 vs. 10 and 5 for FRAX and Garvan, respectively) of risk factors expectedly determines a better discrimination for Qfractures score, it also leads to complex implementation and adherence issues. Both Garvan and QFracture have never been validated using an Italian data source.

The FRAX score therefore encompasses the greatest number of independent studies, was evaluated in the largest number of countries, and is adequately controlled for the background risk of death.[9, 10, 48, 49]

However, the first Italian version of FRAX was developed and updated using administrative data sources, which are based on hospitalization records. The authors estimated an incidence rate of hip/femur fractures of 151 and 399 per 100,000 person-years for men and women, respectively. Concerning the ten-year predicted risk of major osteoporotic fractures, it was 3.4 and 7.5% among men carrying no risk factor and with prior fractures, respectively. In women, it was 5.7 and 11%. Concerning hip/femur fractures, the estimates were 0.8 and 1.9% ten-year predicted risk of hip/femur fractures in men, while they were 1.2 and 3.9% in women.[15]

Unfortunately, FRAX algorithm is a proprietary risk adjustment system which is not transparent and cannot be directly applied to our cohort. Furthermore, the use of claims databases allows the identification of osteoporotic fractures through hospitalization records. This approach provides an accurate definition for hip/femur fractures, while the other osteoporotic fractures, which are not consistently hospitalized, are not sensible enough to be predicted using FRAX accurately. Furthermore, administrative data sources do not comprise empirical information on BMI and smoking, which have been indeed derived from previous studies and/or meta-analyses.

Given the gatekeeper role of GPs in public health systems, we attempted to overcome the aforementioned shortcomings using primary care data. They collect many years of follow-up along with lifestyle and clinical variables, which are dynamically updated over time and are not limited to hospitalization records. Although we dealt with several missing values for BMI and smoking, the distribution of patients among the related categories appear reliable when compared with those reported by previous national surveys. As expected, we found a greater prevalence of obese/overweight patients (30 vs. 10%) and current smokers (40 vs. 22%). This is likely due to characteristics of subjects referring to primary care, who are generally older (i.e. pediatric patients are excluded) and sicker than those targeted by prior surveys.[50, 51]

Overall, the FRA-HS algorithm has an acceptable accuracy, which is mainly due to its performance for hip/femur fractures in women. Instead, the algorithm should be refined for overall major osteoporotic fractures. Indeed, the lower explained variance seen for this latter outcome suggests the presence of unmeasured (i.e. not available in HSD) risk factors able to increase the predictive ability. These results have been confirmed by the explained variance, discrimination and calibration measures obtained in the external validation dataset. Furthermore, the higher predictive accuracy seen for women was consistent with that observed for FRAX® score.[1, 9, 48, 52]

As in previous Italian studies as well as other countries, osteoporotic fractures were more frequent in older patients and particularly in women. Concerning the clinical risk factors, underweight and long-term use of oral corticosteroids showed the highest weights for both femur/hip and overall major osteoporotic fractures. The contribution of these risk factors was greater in females and increased with age. Although most of the observed effect for individual risk factors of FRA-HS were similar to those estimated in the first version (UK) of FRAX,[9] some differences can be due to the different setting and different population. In this respect, it is worth underling that every prediction model needs to be developed and validated in the population in which it is finally applied.[35]

The FRA-HS score has potential limitations as well. First, the fact that we were unable to capture all hospitalization-related fractures likely led to underestimation of femur/hip fractures. However, the reliability of hip/femur fractures diagnoses has been demonstrated previously in HSD,[18] and our estimates showed a little underestimation when compared with those derived from the hospitalization records.[15, 53] Instead, the incidence rates for humeral, vertebral and wrist/forearm fractures which rarely needs hospital admission, were expectedly greater than those obtained from hospital-related data. Second, we did not have information on family history for osteoporotic fractures, which is another clinical risk factor included in the first version of FRAX.[54] However, given the good accuracy of our score for hip/femur fractures, this variable should be above and beyond the other factors already included in the model, while the presence of this variable should be able to improve the predictive accuracy for overall major osteoporotic fractures. Third, the predicted risk showed some over and underestimation when plotted against observed risk. Nevertheless, the confidence intervals for spline-based regression mostly overlapped the line of perfect calibration, so indicating the presence of moderate calibration. This result is realistic given that calibration assessment is highly dependent on the sample size of the dataset, and the presence of interaction terms in the model.[33] Furthermore, the score has

better accuracy for patients belonging to the lower risk categories (i.e., low-moderate risk) of FRA-HS, who expectedly represent most of the patients being cared in the primary care setting. Instead, those categories with greater risk of fractures showed more unstable estimates (underestimation of risk as shown in Figure 1) because of the reduced sample size. However, from a clinical perspective, it is hard to imagine that GPs are not going to treat high-risk patients. Finally, the presence of several missing data for BMI and smoking might have biased the results. However, when we imputed missing values in the models, the results were generally consistent with those yielded for the primary models. Furthermore, although the exclusion of patients with missing values (Model 3 and Model 4) entailed little improvement in calibration, we found lower discrimination than that revealed by primary models, and some risk factors were counterintuitively protective (see Supplementary Table 1). Along this line, model with imputed missing data showed an unexpected reduced risk for smokers. These results might indicate that data on patients with no missing data (i.e. likely those featured by greater baseline risk) are likely better registered than those related to patients with an apparent lower risk, so suggesting the presence of Berksonian bias.[55] Also for this reason, the primary model on femur/hip fractures appears more suitable for potential application in clinical practice.

In conclusion, the FRA-HS has an acceptable accuracy for the prediction of hip/femur fractures in both men and women, while there is room for improvement in prediction model for overall major osteoporotic fractures. These findings would allow the FRA-HS implementation in Italian GPs' software for automatic calculation of risk of hip/femur fracture, so guiding the clinical decision making process and tailoring patient's management and therapy.

**Funding:** This study has been supported by the Italian College of General Practitioners and Primary Care.

**Conflict of Interest:** Dr. Francesco Lapi, Dr. Iacopo Cricelli provided consultancies in protocol preparation for epidemiological studies and data analyses for Servier and Abiogen. Dr. Claudio Cricelli provided clinical consultancies for Servier and Abiogen. Dr. Elisa Bianchini, Dr. Raffaella Michieli, Dr. Alessandro Pasqua, Dr. Giampiero Mazzaglia, Dr. Daniel Prieto-Alhambra, Prof. Bruno Frediani, and Prof. Maria Luisa Brandi have no conflict of interests to disclose.

## References

1. Cooper C, Cole ZA, Holroyd CR, et al (2011) Secular trends in the incidence of hip and other osteoporotic fractures. *Osteoporos Int* 22:1277–1288. doi: 10.1007/s00198-011-1601-6
2. Cauley JA, Chalhoub D, Kassem AM, Fuleihan GE-H (2014) Geographic and ethnic disparities in osteoporotic fractures. *Nat Rev Endocrinol* 10:338–351. doi: 10.1038/nrendo.2014.51
3. Watts NB (2014) Insights from the Global Longitudinal Study of Osteoporosis in Women (GLOW). *Nat Rev Endocrinol* 10:412–422. doi: 10.1038/nrendo.2014.55
4. Johnell O, Kanis JA, Oden A, et al (2004) Mortality after osteoporotic fractures. *Osteoporos Int* 15:38–42. doi: 10.1007/s00198-003-1490-4
5. Johansson H, Oden A, Kanis J, et al (2010) Low bone mineral density is associated with increased mortality in elderly men: MrOS Sweden. *Osteoporos Int*. doi: 10.1007/s00198-010-1331-1
6. Cannada LK, Hill BW (2014) Osteoporotic Hip and Spine Fractures: A Current Review. *Geriatr Orthop Surg Rehabil* 5:207–212. doi: 10.1177/2151458514548579
7. Ratti C, Vulcano E, La Barbera G, et al (2013) The incidence of fragility fractures in Italy. *Aging Clin Exp Res* 25 Suppl 1:S13-4. doi: 10.1007/s40520-013-0071-4
8. Rubin KH, Abrahamsen B, Friis-Holmberg T, et al (2013) Comparison of different screening tools (FRAX(R), OST, ORAI, OSIRIS, SCORE and age alone) to identify women with increased risk of fracture. A population-based prospective study. *Bone* 56:16–22. doi: 10.1016/j.bone.2013.05.002
9. Kanis JA, Johnell O, Oden A, et al (2008) FRAX and the assessment of fracture probability in men and women from the UK. *Osteoporos Int* 19:385–397. doi: 10.1007/s00198-007-0543-5
10. Marques A, Ferreira RJO, Santos E, et al (2015) The accuracy of osteoporotic fracture risk prediction tools: a systematic review and meta-analysis. *Ann Rheum Dis*. doi: 10.1136/annrheumdis-2015-207907
11. Collins GS, Le Manach Y (2015) Flawed validation of FRAX. *Bone*. doi: 10.1016/j.bone.2015.07.005
12. Kanis JA, Johnell O, De Laet C, et al (2002) International variations in hip fracture probabilities: implications for risk assessment. *J Bone Miner Res* 17:1237–1244. doi: 10.1359/jbmr.2002.17.7.1237
13. Kanis JA, Hans D, Cooper C, et al (2011) Interpretation and use of FRAX in clinical practice. *Osteoporos Int* 22:2395–411. doi: 10.1007/s00198-011-1713-z
14. Kreidieh OI, El-Hajj Fuleihan G (2014) Impact of changes in mortality on FRAX-derived fracture probabilities. *Bone* 62:43–50. doi: 10.1016/j.bone.2014.01.014
15. Piscitelli P, Chitano G, Johannson H, et al (2013) Updated fracture incidence rates for the Italian version of FRAX(R). *Osteoporos Int* 24:859–866. doi: 10.1007/s00198-012-2021-y
16. Oden A, Dawson A, Dere W, et al (1998) Lifetime risk of hip fractures is underestimated.

17. Bianchini E, Brignoli O, Cricelli C, et al (2014) VIII Report Health Search. Anno 2013/2014. Firenze
18. Lapi F, Simonetti M, Michieli R, et al (2012) Assessing 5-year incidence rates and determinants of osteoporotic fractures in primary care. *Bone* 50:85–90. doi: 10.1016/j.bone.2011.09.048
19. Sterrantino C, Trifiro G, Lapi F, et al (2013) Burden of community-acquired pneumonia in Italian general practice. *Eur Respir J* 42:1739–1742. doi: 10.1183/09031936.00128713
20. Mazzaglia G, Ambrosioni E, Alacqua M, et al (2009) Adherence to antihypertensive medications and cardiovascular morbidity among newly diagnosed hypertensive patients. *Circulation* 120:1598–605. doi: 10.1161/CIRCULATIONAHA.108.830299
21. (2016) Mille in Rete [http://www.svemg.it/?page\\_id=1783](http://www.svemg.it/?page_id=1783). [http://www.svemg.it/?page\\_id=1783](http://www.svemg.it/?page_id=1783). Accessed 28 Oct 2016
22. Leslie WD, Lix LM (2014) Comparison between various fracture risk assessment tools. *Osteoporos Int* 25:1–21. doi: 10.1007/s00198-013-2409-3
23. Steyerberg EW, Vickers AJ, Cook NR, et al (2010) Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 21:128–38. doi: 10.1097/EDE.0b013e3181c30fb2
24. Steyerberg EW, Bleeker SE, Moll HA, et al (2003) Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *J Clin Epidemiol* 56:441–447. doi: 10.1016/S0895-4356(03)00047-7
25. Melton LJ, Thamer M, Ray NF, et al (1997) Fractures attributable to osteoporosis: report from the National Osteoporosis Foundation. *J Bone Miner Res* 12:16–23. doi: 10.1359/jbmr.1997.12.1.16
26. Warriner AH, Patkar NM, Curtis JR, et al (2011) Which fractures are most attributable to osteoporosis? *J Clin Epidemiol* 64:46–53. doi: 10.1016/j.jclinepi.2010.07.007
27. Curtis JR, Taylor AJ, Matthews RS, et al (2009) “Pathologic” fractures: should these be included in epidemiologic studies of osteoporotic fractures? *Osteoporos Int* 20:1969–72. doi: 10.1007/s00198-009-0840-2
28. Ray WA, Griffin MR, Fought RL, Adams ML (1992) Identification of fractures from computerized Medicare files. *J Clin Epidemiol* 45:703–714.
29. Kanis JA, McCloskey E V, Johansson H, et al (2010) Development and use of FRAX in osteoporosis. *Osteoporos Int* 21 Suppl 2:S407–13. doi: 10.1007/s00198-010-1253-y
30. Steinbuch M, Youket TE, Cohen S (2004) Oral glucocorticoid use is associated with an increased risk of fracture. *Osteoporos Int* 15:323–8. doi: 10.1007/s00198-003-1548-3
31. Fine, JP, Gray R (1999) A Proportional Hazards Model for the Subdistribution of a Competing Risk. *J Am Stat Assoc* 496–509.
32. Wolbers M, Blanche P, Koller MT, et al (2014) Concordance for prognostic models with competing risks. *Biostatistics* 15:526–39. doi: 10.1093/biostatistics/kxt059
33. Van Calster B, Nieboer D, Vergouwe Y, et al (2016) A calibration hierarchy for risk



models was defined: from utopia to empirical data. *J Clin Epidemiol* 74:167–76. doi: 10.1016/j.jclinepi.2015.12.005

34. Van Hoorde K, Van Huffel S, Timmerman D, et al (2015) A spline-based tool to assess and visualize the calibration of multiclass risk predictions. *J Biomed Inform* 54:283–93. doi: 10.1016/j.jbi.2014.12.016
35. Hippisley-Cox J, Coupland C (2012) Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study. *BMJ* 344:e3427.
36. Janssen KJM, Vergouwe Y, Donders ART, et al (2009) Dealing with missing predictor values when applying clinical prediction models. *Clin Chem* 55:994–1001. doi: 10.1373/clinchem.2008.115345
37. Knol MJ, Janssen KJM, Donders ART, et al (2010) Unpredictable bias when using the missing indicator method or complete case analysis for missing confounder values: an empirical example. *J Clin Epidemiol* 63:728–36. doi: 10.1016/j.jclinepi.2009.08.028
38. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM (2006) Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 59:1087–91. doi: 10.1016/j.jclinepi.2006.01.014
39. Moons KGM, Donders RART, Stijnen T, Harrell FE (2006) Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol* 59:1092–101. doi: 10.1016/j.jclinepi.2006.01.009
40. Collins GS, Reitsma JB, Altman DG, Moons KGM (2015) Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *J Clin Epidemiol* 68:134–43. doi: 10.1016/j.jclinepi.2014.11.010
41. van der Heijden GJMG, Donders ART, Stijnen T, Moons KGM (2006) Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol* 59:1102–9. doi: 10.1016/j.jclinepi.2006.01.015
42. Sterne JAC, White IR, Carlin JB, et al (2009) Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 338:b2393.
43. Moons KGM, Donders RART, Stijnen T, Harrell FE (2006) Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol* 59:1092–101. doi: 10.1016/j.jclinepi.2006.01.009
44. Kanis JA, Johansson H, Oden A, et al (2014) Worldwide uptake of FRAX. *Arch Osteoporos* 9:166. doi: 10.1007/s11657-013-0166-8
45. Azagra R, Roca G, Encabo G, et al (2012) FRAX(R) tool, the WHO algorithm to predict osteoporotic fractures: the first analysis of its discriminative and predictive ability in the Spanish FRIDEX cohort. *BMC Musculoskelet Disord* 13:204. doi: 10.1186/1471-2474-13-204
46. Blanche P, Dartigues J-F, Jacqmin-Gadda H (2013) Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* 32:5381–97. doi: 10.1002/sim.5958
47. Kanis JA, Johnell O, Oden A, et al (2000) Long-term risk of osteoporotic fracture in

Malmo. Osteoporos Int 11:669–674.

48. McCloskey E V, Johansson H, Oden A, Kanis JA (2009) From relative risk to absolute fracture risk calculation: the FRAX algorithm. *Curr Osteoporos Rep* 7:77–83.
49. WD L, LM L, Johansson H, et al (2010) Independent clinical validation of a Canadian FRAX((R)) tool: Fracture prediction and model calibration. *J Bone Miner Res*. doi: 10.1002/jbmr.123
50. <http://www.epicentro.iss.it/temi/fumo/EpidItalia.asp>.
51. <http://www.epicentro.iss.it/problemi/obesita/EpidItalia.asp>.
52. Couris CM, Chapurlat RD, Kanis JA, et al (2012) FRAX® probabilities and risk of major osteoporotic fracture in France. *Osteoporos Int* 23:2321–7. doi: 10.1007/s00198-011-1883-8
53. Piscitelli P, Tarantino U, Chitano G, et al (2011) Updated incidence rates of fragility fractures in Italy: extension study 2002-2008. *Clin Cases Miner Bone Metab* 8:54–61.
54. McGowan B, Kanis JA, Johansson H, et al (2013) Development and application of FRAX in the management of osteoporosis in Ireland. *Arch Osteoporos* 8:146. doi: 10.1007/s11657-013-0146-z
55. Westreich D (2012) Berkson’s bias, selection bias, and missing data. *Epidemiology* 23:159–64. doi: 10.1097/EDE.0b013e31823b6296

## Figures captions

**Fig. 1** Flexible calibration curve plotting predicted vs. observed risk of overall major osteoporotic fractures.

**Fig. 1 legend:** CI: Confidence Intervals; spikes above the x-axis are cases of fractures, non-cases are below the x-axis.

**Fig. 2** Ten-year risk of hip/femur fractures (Model 1) or overall major osteoporotic fractures (Model 2) in the presence or not of individual risk factor among patients aged 60-70 years with normal weight broken down by sex.

**Fig. 3** Ten-year risk of hip/femur fractures (Model 3) or overall major osteoporotic fractures (Model 4) in the presence or not of individual risk factor among patients aged 60-70 years with normal weight broken down by sex: patients with missing data are excluded.

**Supplemental Fig. 1** Performance of multiple imputation method on BMI and smoking: comparison between full dataset and the five imputed datasets.