PROOF COVER SHEET

Author(s): Wojciech Pluskiewicz, Piotr Adamczyk, Edward Franek, Ewa Sewerynek, Piotr Leszczynski, Hanna Wichrowska, Luiza Napiórkowska, Micha Stuss, Aleksandra Ptaszek, Tomasz Kostyk, Krzysztof S. Golba, Wioleta Garbacz, and Bogna Drozdzowska

Article title: FRAX calculator and Garvan nomogram in male osteoporotic population

Article no: DAGM_A_875991

Enclosures: 1) Query sheet
2) Article proofs

Dear Author,
Please check these proofs carefully. It is the responsibility of the corresponding author to check against the original manuscript and approve or amend these proofs. A second proof is not normally provided. Informa Healthcare cannot be held responsible for uncorrected errors, even if introduced during the composition process. The journal reserves the right to charge for excessive author alterations, or for changes requested after the proofing stage has concluded.

The following queries have arisen during the editing of your manuscript and are marked in the margins of the proofs. Unless advised otherwise, submit all corrections using the CATS online correction form. Once you have added all your corrections, please ensure you press the ''Submit All Corrections'' button.

Please review the table of contributors below and confirm that the first and last names are structured correctly and that the authors are listed in the correct order of contribution.

Contrib. No.	Prefix	Given name(s)	Surname	Suffix
1		Wojciech	Pluskiewicz	
2		Piotr	Adamczyk	
3		Edward	Franek	
4		Ewa	Sewerynek	
5		Piotr	Leszczynski	
6		Hanna	Wichrowska	
7		Luiza	Napiórkowska	
8		Micha	Stuss	
9		Aleksandra	Ptaszek	
10		Tomasz	Kostyk	
11		Krzysztof S.	Golba	
12		Wioleta	Garbacz	
13		Bogna	Drozdzowska	

AUTHOR QUERIES

Q1: Please provide the town and state abbreviation for StatSoft, Inc., USA cited in text.

Q2: Please provide the town and country of origin (for other countries) identifying the headquarter location for MedCalc, Belgium cited in text.

Q3: Please check and confirm Author name "Micha? Stuss"

Q4: Please provide better quality artworks for figures 1, 2, 3 and 5.

The Aging Male

http://informahealthcare.com/tam
ISSN: 1368-5538 (print), 1473-0790 (electronic)

Aging Male, Early Online: 1–9
 DOI: 10.3109/13685538.2013.875991

informa healthcare

ORIGINAL ARTICLE

FRAX calculator and Garvan nomogram in male osteoporotic population

Wojciech Pluskiewicz[1], Piotr Adamczyk[2], Edward Franek[3], Ewa Sewerynek[4], Piotr Leszczynski[5,6], Hanna Wichrowska[3], Luiza Napiórkowska[3], Michał Stuss[4], Aleksandra Ptaszek[4], Tomasz Kostyk[5], Krzysztof S. Golba[7], Wioleta Garbacz[8], and Bogna Drozdzowska[9]

[1]*Department and Clinic of Internal Diseases, Diabetology and Nephrology, Metabolic Bone Diseases Unit, Medical University of Silesia, Katowice, Poland,* [2]*Department and Clinic of Paediatrics, Medical University of Silesia, Katowice, Poland,* [3]*Department of Endocrinology, Medical Research Centre, Polish Academy of Science, Warsaw, Poland,* [4]*Department of Endocrine Disorders and Bone Metabolism, Medical University, Lodz, Poland,* [5]*Department of Rheumatology and Osteoporosis, Hospital J. Strusia, Poznan, Poland,* [6]*Department of Physiotherapy, Rheumatology and Rehabilitation, Medical University, Poznan, Poland,* [7]*Department of Cardiology, Medical University of Silesia, Katowice, Poland,* [8]*Military Hospital, Gliwice, Poland, and* [9]*Department of Pathomorphology, Medical University of Silesia, Katowice, Poland*

Abstract

Purpose: The aim of the study was the presentation of osteoporotic fracture prediction in men.
Methods: Eight-hundred and one men at the mean age of 70.8 ± 9.31 years were examined. The 10-year fracture prediction was established, using the FRAX™ calculator and Garvan nomogram.
Results: The mean value for any fracture and hip fracture probabilities for FRAX were 7.26 ± 5.4% and 3.68 ± 4.25%, respectively. For Garvan fracture, risk values were 26.44 ± 23.83% and 12.02 ± 18.1%. The mean conformity for any fracture and hip fracture prediction for threshold of 20% (any fracture) and 3% (hip fracture) between Garvan and FRAX values was 55.8% (κ 0.041) and 79.65% (κ 0.599), respectively. ROC analyses showed the following areas under the ROC curves (AUC) for any fractures: FRAX 0.808 and Garvan nomogram 0.843 ($p = 0.059$). The AUC values for hip fractures were 0.748 for Garvan nomogram and for 0.749 FRAX, and did not differ. On the base of ROC data, the cut-off values with best accuracy to predict fractures for both methods were established. The conformity between methods for thresholds indicated by ROC analysis was 72.5% (κ 0.435) for any and 77.7% (κ 0.543) for hip fractures.
Conclusion: The conformities between FRAX and Garvan in regard to hip fracture prediction were acceptable for a threshold of 3% and thresholds derived by ROC analysis, while for any fracture we recommend to use thresholds established by ROC analysis. This may suggest that the use of "universal" cut-off points is probably misleading.

Keywords

Fracture probability, FRAX, garvan nomogram, men, osteoporotic fracture

History

Received 7 July 2013
Revised 17 September 2013
Accepted 6 December 2013
Published online ▮▮▮

Introduction

Osteoporosis, a serious health problem, is commonly considered as women's disease. The majority of patients are women but, especially at later age stages, a significant part of all osteoporotic patients are men. The lifetime fracture risk in male subjects after the age of 40 amounts to 25% [1] and it is the men who are more frequently affected by serious consequences of experienced hip fracture(s). Also, the knowledge on bone metabolism and the awareness of osteoporosis as such are much weaker in men than in women. Osteoporosis is usually a clinically silent disease and, fairly often, an osteoporotic fracture comes up as its first manifestation. As it is widely known, fracture history is one of the strongest risk factors for subsequent fractures [2], therefore, the main goal of any osteoporosis management is primary prevention of osteoporotic fractures. In order to properly set up prophylactic therapy, an accurate assessment of fracture probability or risk is an imperative. Recently, some prognostic models have been developed [3–6]. They are based on bone density measurements and take into consideration several defined clinical risk factors. In 2008, the WHO introduced a new fracture prediction tool (FRAX algorithm) to determine patient's absolute fracture probability over a 10-year span [7]. The FRAX algorithm was developed by the WHO to be applicable in men and postmenopausal women; the National Osteoporosis Foundation (NOF) Clinician's Guide focuses on its utility in postmenopausal women and men, aged >50. The current NOF Guide recommends to examine patients, taking into account their 10-year, FRAX-estimated probability scores of ≥3% for hip fracture and ≥20% for major osteoporotic

Address for correspondence: Prof. Wojciech Pluskiewicz, Department and Clinic of Internal Diseases, Diabetology and Nephrology – Metabolic Bone Diseases Unit in Zabrze, Medical University of Silesia in Katowice, 3-Maja 13/15 street, 41-800 Zabrze, Poland. Tel/Fax: +48 323704389. E-mail: osteolesna@poczta.onet.pl

fracture, in order to reduce their general fracture risk [8]. In turn, other authors have proposed algorithms for individualized 5-year and 10-year fracture risk prognoses, applicable for both women and men [9,10]. Obviously, neither fracture risk nor fracture probability assessment can entirely replace objective examination by a doctor, and only an absolute and comprehensive fracture risk prediction should be accounted for an appropriate management approach. The majority of published reports on fracture risk or probability address women, while a few papers only report studies (also scarce in their number) conducted in men [11–14].

The purpose of the reported cross-sectional study was to present and validate a 10-year fracture prediction in a group of 801 Polish men, determined by both the FRAX calculator [7] and the nomogram, proposed by Nguyen et al. from the Garvan Institute [9,10]. Although these two calculators conform each other with essential differences in their design (FRAX takes into account predicted lifetime, whereas Garvan algorithm is not adjusted for the risk of death in aging patients), from the clinical point of view they serve the same purpose, thus their comparison seems to be a justified and important goal.

Material

The studied group included 801 men at the mean age of 70.8 ± 9.31 years (the age range 55–94 years), evaluated at four osteoporotic outpatient clinics in four different centers. The entire study group comprised all successive patients, attending the clinics during the period from May 2009 to June 2010. The mean parameter values and SDs for weight, height and BMI were: 78.3 ± 13.2 kg, 169.7 ± 7.0 cm and 27.2 ± 3.91 kg/m^2, respectively.

A group of 218 men (27.2%) had, at least, one low-traumatic fracture at the age above 45, and 73 men (9.1%) presented with the history of multiple fractures, amounting to a total of 371 fractures in both groups. Therefore, a subgroup of 218 men include men with only one fracture ($n = 145$) and those who had more than one fracture ($n = 73$). The fractures were identified in the following skeletal sites: spine ($n = 206$), distal forearm ($n = 70$), tibia or fibula ($n = 39$), ribs ($n = 31$), proximal femur ($n = 16$), humerus ($n = 9$). Generally, the diagnosis of fracture was based on patients medical documentation including X-ray but in some patients spine fractures were self-reported because X-ray were not available. In patients who presented spine radiograms at the moment of recruitment, vertebral fractures were diagnosed according to widely accepted rules proposed by Genant.

The total number of patients with other clinical risk factors for osteoporosis, taken into consideration for fracture prediction in the studied cohort men, included those with hip fracture history in parents ($n = 46$, 5.7%), those on steroid therapy ($n = 82$, 10.3%), subjects with rheumatoid arthritis ($n = 35$, 4.4%), with secondary osteoporosis ($n = 43$, 5.8%), alcohol abuse ($n = 24$, 3.0%) and – finally – the number of patients with falls (one or more) during the last 12 months ($n = 90$, 11.2%).

Methods

Fracture prediction was assessed by the FRAX [5] calculator (http://www.shef.ac.uk/FRAX) and the Garvan nomogram [6,7] (http://www.garvan.org.au/bone-fracture-risk). The 10-year fracture probability by the FRAX algorithm was based on age, BMI, fracture history in adulthood, hip fracture in parents' history, steroid use, rheumatoid arthritis, alcohol abuse, secondary osteoporosis and T-scores for femoral neck BMD (Bone Mineral Density). Fracture history was determined from patient reports and only fall fractures from standing height (an example typical for osteoporosis) were taken into consideration. T-scores, used for the calculations, were derived from NHANES III database for young females in all DXA devices were used. In order to calculate fracture probability by the FRAX algorithm, a model for the Polish male population was applied.

The 10-year fracture risk, estimated by Garvan nomogram, was based on the age, the number of prior fractures after 50, the number of falls during previous 12 months and T-score values for femoral neck BMD.

The FRAX calculator produces estimates for ''major fractures", in general confined to hip, humerus, spine and wrist, whereas the Garvan ''all fractures'' category is much broader and includes more fracture sites. This methodological difference was very important for the interpretation of results in our comparative study. As a rule, radiography was not used to confirm fracture occurrence.

A low fracture risk/probability for any fracture was defined when the value was $<20\%$, while high fracture risk/probability was accepted in cases of $\geq 20\%$. The respective values for hip fracture risk/probability were $<3\%$ and $\geq 3\%$.

The data for evaluation were acquired from bone densitometry centres in four Polish cities (Zabrze, Lodz, Warsaw and Poznan), covering the period from May 2009 to June 2010. In order to collect necessary data for fracture prediction, a structured questionnaire was used and the data were collected by physicians. All the subjects were submitted to hip BMD [g/cm^2] measurements. Three GE Lunar and one Norland densitometer were used for that purpose.

The reported study received an approval of the local ethics committee.

Statistics

Statistical analysis was performed by means of the Microsoft Office Excel application, the Statistica 8 program (StatSoft, Inc., USA) and MedCalc 11.1.1.0 (MedCalc, Belgium), run on a PC computer. Fracture prediction was calculated for each studied subject, according to the FRAX algorithm [5] and given by Garvan nomogram [6,7]. Descriptive statistics are presented as the mean values and standard deviations (SDs). In order to apply an analysis of conformity assessment by both methods, the studied group was divided in the following ways [8]:

- two thresholds (levels) of fracture risk (Garvan) or probability (FRAX) in case of any fracture (low risk $<20\%$ and high risk $\geq 20\%$),
- two thresholds (levels) of fracture risk (Garvan) or probability (FRAX) in case of hip fracture (low risk $<3\%$ and high risk $\geq 3\%$).

The conformity was established separately for any fracture risk and for hip fracture risk, being defined as the same fracture risk threshold in either method (either low fracture

risk or high fracture risk for both calculators in particular patient). A reverse situation (low fracture risk, according to one method, and high fracture risk by the other) was classified as disconformity. The receiver-operating characteristics (ROC) curve analysis was applied: (1) to compare the diagnostic performance of the FRAX algorithm and Garvan nomogram in the assessment of any and hip fracture prediction and (2) to set a decision-making cut-off value of risk/probability which corresponded to the optimal threshold point of the ROC curve determined by Youden index. The calculation of the AUCs in ROC analysis has been done based on the dichotomous variable of fracture. As a negative case the patient without any prior fracture has been used, while as a positive case a patient with either any fracture or with hip fracture.

We also performed an additional analysis of conformity, using cut-off values, determined on the basis of the Youden index, established by ROC analysis for hip and any fracture instead of *a priori* assumed cut-off points at 3% and 20%.

For further analysis, it was assumed that the FRAX algorithm would be regarded as the reference method of fracture probability and the index of conformity was calculated as the percentage of men, classified at a given, FRAX-established risk level, who achieved the same risk threshold values in Garvan nomogram. The level of conformity (agreement) was presented as the percentage values and as the results of Cohen's κ test. The Wilcoxon test was used to compare the results of conformity expressed in percentage values before and after the correction based on ROC analysis.

A comparison of fracture predictive values was performed with the Mann–Whitney's U-test in subgroups with and without fracture history. All the results of statistical tests were regarded as statistically significant, when $p < 0.05$.

Results

Fracture probability or risk

The mean value for any and hip fracture probabilities in FRAX were $7.26 \pm 5.4\%$ and $3.68 \pm 4.25\%$, respectively. In Garvan estimates, the fracture risk values were $26.44 \pm 23.83\%$ and $12.02 \pm 18.1\%$. Table 1 presents the numbers of men with low and high fracture predictive values for FRAX and Garvan nomogram.

Figure 1 shows fracture prediction changes with advancing age. The values from Garvan nomogram for any fracture were higher than those, obtained by the FRAX algorithm. However, that difference was not significant till the age of about 60. In hip fracture assessments, higher values were maintained till the age of 65, following the FRAX model. Then, with advancing age, the Garvan model values were higher. The average values of fracture probability or risk, derived for the whole study group, are presented in Figure 2.

Conformity with 20% threshold – any fracture

The mean conformity for any fracture prediction between Garvan and FRAX values was 55.8%, which corresponds to Cohen's κ value of 0.041, reflecting very poor level of agreement between methods. Of the 801 men, 428 were classified by both methods as low risk/probability and 19 were classified by both methods as high risk/probability. 348 demonstrated high risk, according to the Garvan method, and low probability, according to FRAX. Six were estimated with low Garvan and high FRAX values. These data are shown in Figure 3, part A.

Table 1. Subjects with low and high fracture probability by the FRAX calculator and those with low and high fracture risk by the Garvan algorithm.

Applied model	Number of men with low probability/risk	Number of men with high probability/risk
Any fracture probability by FRAX	776	25
Hip fracture probability by FRAX	451	350
Any fracture risk by Garvan	434	367
Hip fracture risk by Garvan	356	445

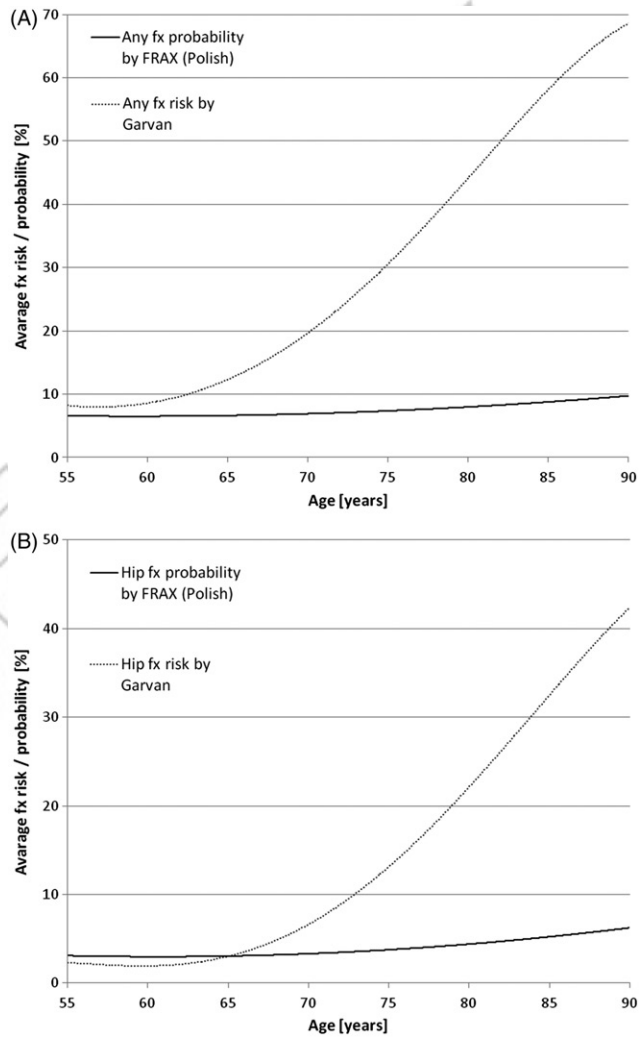

Figure 1. Changes in any fracture risk/probability (A) and hip fracture risk/probability (B) over age range. The curves demonstrate trends of the relations between calculated fracture risk/probability and age in individual subjects of the study cohort.

Conformity with 3% threshold – hip fracture

The mean conformity for hip fracture prediction between Garvan and FRAX values was 79.65%, corresponding to

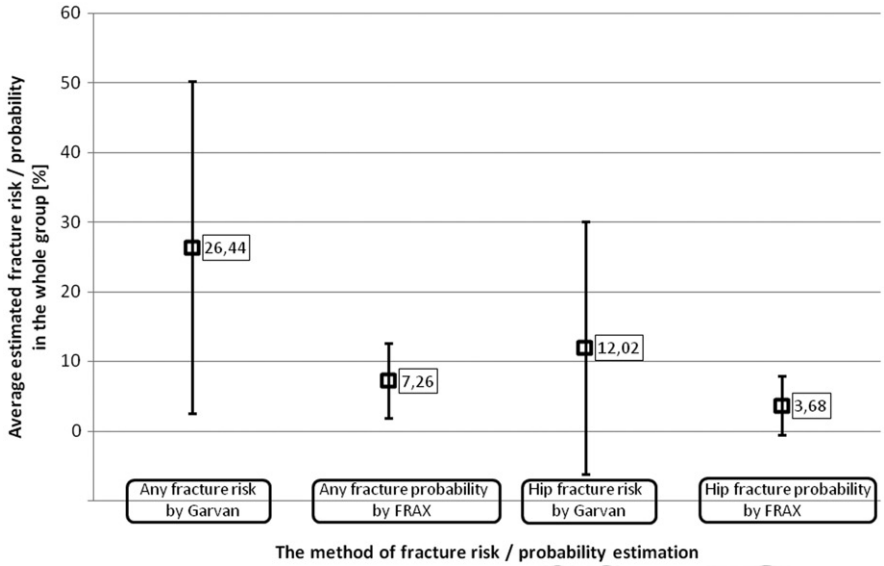

Figure 2. The average values of fracture probability (for FRAX) or fracture risk (for Garvan) derived from the whole study group.

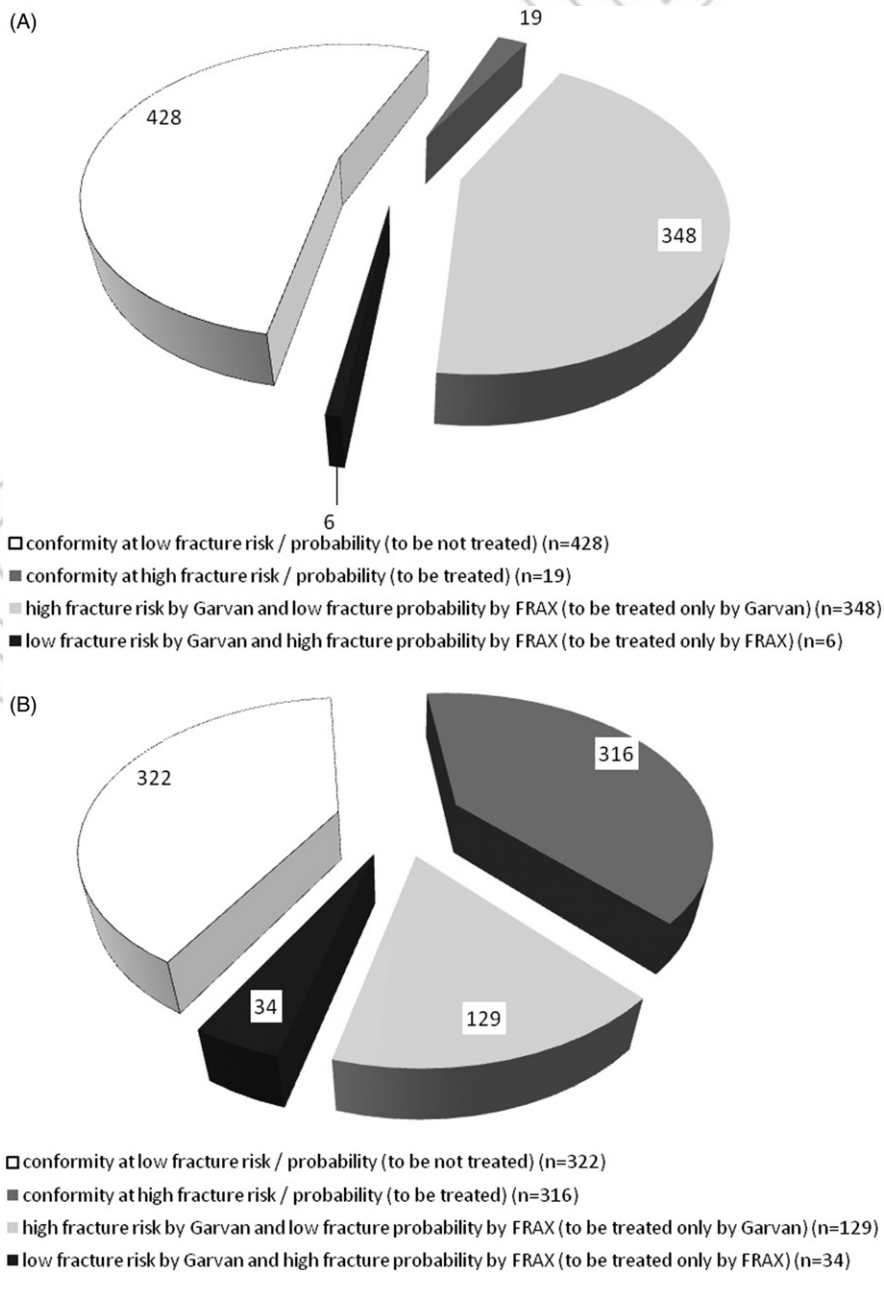

Figure 3. Subgroups of men with or without indications for treatment, according to FRAX, based on the Polish reference population, and Garvan, established from ''routine'' cut-off values (20% in any fracture assessment and 3% in hip fracture assessment) for any fracture risk (A) or for hip fracture risk (B).

Cohen's κ value of 0.599. So the agreement between methods for hip fracture was much better than that for any fractures and could be classified at ''moderate'' (even close to ''good'') agreement level. Of the 801 men, 322 were classified by both methods at low risk/probability level, 316 were classified by both methods at high risk/probability level. 129 demonstrated high risk, according to Garvan and low probability, according to FRAX. In 34, low Garvan and high FRAX values were noted. These data are shown in Figure 3, part B.

The most important factors, resulting in different classification levels (high Garvan and low FRAX values), included falls and multiple fractures, and the observed opposite situation was associated with a high number of clinical risk factors, especially smoking, steroid use and rheumatoid arthritis.

Fracture risk/probability stratified by fracture status

Table 2 presents fracture prediction levels, according to fracture status, e.g. the presence or absence of fracture. For both methods and for any and hip fracture, the fracture risk or probability was significantly higher in men with fracture(s) in history, in comparison with those without such medical records. However, the probability levels, estimated by both FRAX algorithms, approximately doubled in the men with fractures in history, in comparison to those without previous fracture episodes. The fracture risk level, according to Garvan nomogram, was three times higher. One can read from Table 2 that previous fracture remains the strongest factor, influencing the risk/probability of consecutive fracture.

Indication for treatment

Traditionally, the initiation of pharmacological treatment is based mainly on T-score and/or the presence of typical osteoporotic fracture. More recently, a 10-year fracture probability or risk was proposed as the method of qualification to treatment, and a threshold of 20% and 3% for any and hip fracture, respectively, was recommended. We analyzed the data to find out how many patients with the traditional indication for treatment presented with high 10-year fracture risk/probability.

Low T-score as an indication to treatment – any fracture

In 251 men (31.33%), T-score value for femoral neck BMD was equal or below −2.5. Among 251 men with T-score below −2.5, only 24 revealed FRAX value above 20%. That means that a significant majority of group studied ($n = 227$, 90.44%) would not be treated, despite low BMD values, due low FRAX values. Opposite results are presented from the Garvan algorithm and only 69 men (27.5%) were classified in low fracture risk, despite low BMD levels, and the majority of those 251 men ($n = 182$; 72,5%) achieved high fracture risk score (the classification of men with low T-score (below −2,5) at high fracture risk category more frequent according to the Garvan algorithm in comparison to FRAX; $\chi^2 = 205.5$, df = 1, $p < 0.0001$).

Low T-score as an indication for treatment – hip fracture

An analysis, regarding hip fracture prediction, showed a much better conformity for both methods. In FRAX, only 18 subjects (7.17%) with low BMD were classified into the group with low fracture risk (<3%). In turn, Garvan nomogram revealed 22 men (8.76%) not qualifying to treatment despite low BMD values (no significant difference between the Garvan algorithm and FRAX calculator for proportion of men with low and high fracture risk among patients with low T-score; $\chi^2 = 0.43$, df = 1, $p < 0.51$).

Prior fracture as an indication to treatment – any fracture

Among 218 men with prior fracture, only 17 (7.8%) presented with the FRAX value above 20%, which means that a significant majority of the patients ($n = 201$, 92.2%) would not be treated, despite positive fracture history, due to low FRAX values. Opposite results are presented by the Garvan algorithm, by which only 39 men (17.9%) were classified at low fracture risk level despite prior fracture and the majority ($n = 179$; 82.1%) achieved high fracture risk score (the classification of men with positive previous fracture history at high fracture risk category more frequent according to the Garvan algorithm in comparison to FRAX; $\chi^2 = 243.3$, df = 1, $p < 0.0001$).

Prior fracture as an indication for treatment – hip fracture

Analogous analysis for hip fracture cases showed better conformity for both methods. In FRAX, 63 subjects (28.9%) with fracture history were classified at low fracture risk (<3%). In Garvan nomogram, only 43 men (19.72%) would be not treated despite prior fracture, whereas 175 (80.3%) achieved high fracture risk score (in comparison to 155 (71.1%) by FRAX). Nonetheless, the differences between methods remains significant as the classification of men with positive previous fracture history at high fracture risk category is slightly more frequent according to the Garvan algorithm in contrast to FRAX; $\chi^2 = 4.99$, df = 1, $p < 0.05$).

Table 2. Assessment of fracture prediction stratified by fracture status.

Applied model	Men without fracture(s) in history ($n = 583$)	Men with fracture(s) in history ($n = 218$)	p
Any fracture probability by FRAX	5.75 ± 3.87%	11.28 ± 6.7%	<0.000001
Hip fracture probability by FRAX	2.77 ± 3.07%	6.11 ± 4.8%	<0.000001
Any fracture risk by Garvan	17.82 ± 14.7%	49.48 ± 27.92%	<0.000001
Hip fracture risk by Garvan	7.7 ± 12.66%	23.62 ± 24.45%	<0.000001

ROC analyses

In order to establish the accuracy of both methods, ROC analyses were performed for any and hip fractures. The statistical significance of the differences between the areas under ROC curves was assessed, using the method of DeLong et al. [15]. The area under ROC curve (AUC) for any fractures is presented in Figure 4, part A. The AUCs from ROC curve analysis were as follows: for the FRAX algorithm – 0.808 (95%CI 0.779–0.835), and for Garvan nomogram – 0.843 (95%CI 0.815–0.867). A borderline significant difference was observed between the obtained AUCs ($p = 0.059$). The area under ROC curve for hip fractures is presented in Figure 4, part B. The AUCs from the ROC curve analysis were as follows: for the FRAX algorithm – 0.749 (95% CI 0.718–0.779), and for Garvan nomogram – 0.748 (95%CI 0.716–0.778). The AUCs did not differ significantly.

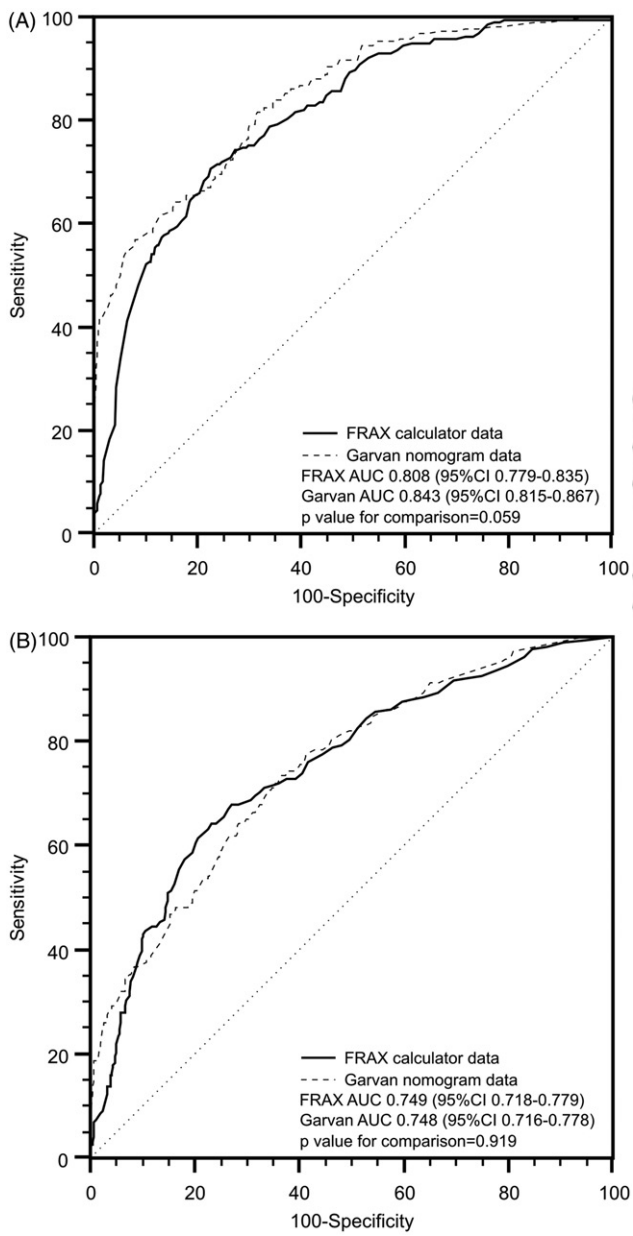

Figure 4. A comparison of ROC (receiver-operating characteristic) curves, drawn for fracture probability, assessed by the FRAX nomogram and for fracture risk assessment by the Garvan nomogram. Part A of the Figure illustrates the probability/risk assessment for any fracture, while Part B for hip fracture.

Based on the ROC curve analysis, new cut-off values were established to predict fractures by each method. The cut-off values for any fracture were as follows: 20.2% (sensitivity, 81.65%; specificity, 68.44%) for the Garvan method and 7.6% (sensitivity, 70.64%; specificity 77.36%) for FRAX. For hip fractures, the respective results were: 4.9% (with the sensitivity of 73.39% and the specificity of 63.46%) and 3.8% (with the sensitivity of 64.22% and the specificity of 76.67%).

Conformity assessment with cut-off values from ROC analysis

Since the ROC analysis prompted other (i.e. at the level of 20% for any fracture and 3% for hip fracture) than routinely accepted cut off points between the patients with low and high fracture risks, we decided to re-establish the conformity between methods, using our own calculated cut-off points (according to the values given in *ROC analyses* section).

When the patients were classified at low or high fracture risk, according to thresholds indicated by ROC analysis (different cut-off points for FRAX and Garvan method), the conformity between the methods was 72.5% (κ 0.435 – moderate agreement between methods) for any fracture and 77.7% (κ 0.543 – also moderate agreement between methods) for hip fracture. Figure 5 presents the numbers of men, classified at low and high fracture risk level, according to both methods and with cut-off points, established in the ROC analysis. When our ROC analysis-based conformity was compared by the Wilcoxon test to conformity assessed by ''routine'' criteria, a significant improvement was observed in the conformity between the methods for any fracture ($p < 0.0001$) and no change for hip fracture conformity.

Discussion

In the reported study, data of a 10-year fracture prediction in men are presented, with regards to some significant clinical points (age-related changes, the conformity in respect to established therapeutic thresholds, the influence of prior fracture(s), ROC analysis), assessed by the FRAX algorithm and Garvan nomogram, providing fairly unexpected results; both methods provided any fracture prediction with poor conformity (agreement) of 55.8% (κ 0.041). In turn, the comparisons for hip fracture prediction, with the value of 79.65% (κ 0.599), demonstrated a fairly good agreement level. However, the use of cut-off values, which corresponded to an optimal threshold point from ROC curve analysis, improved significantly the conformity for any fracture, while the conformity for hip fracture was almost the same. These results suggest that treatment decision should be based rather on the thresholds for fracture prediction, defined for a specific population, instead of fixed criteria.

The conformity level, obtained in the reported study, should be compared with the recent results of a similar analysis, performed in a group of 2012 Polish women [16]. In that study, carried out with the use of the US Caucasian FRAX model and the Garvan algorithm (at that time, the Polish FRAX was not yet available), the conformity level was

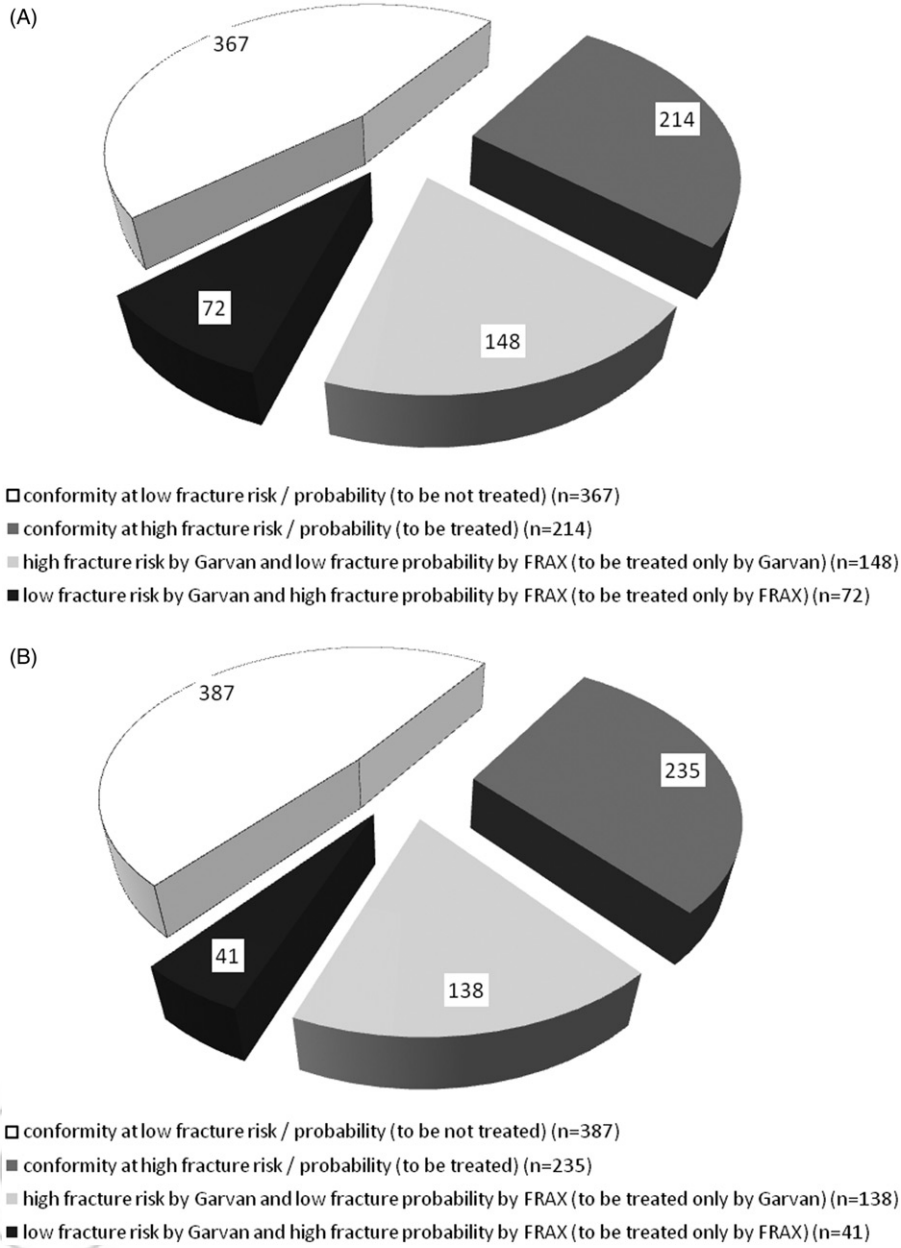

Figure 5. Subgroups of men with or without indications for treatment, according to FRAX, based on the Polish reference population, and by the Garvan method, using the cut-off values, calculated from ROC curves (for details see the text) for any fracture risk (A) or for hip fracture risk (B).

79.1% for any and 79.5% for hip fracture prediction. The currently obtained data for men are much weaker for any fracture and almost the same for hip fracture prediction. In both studies, the values of 3% for hip and 20% for any fracture were used as the thresholds for treatment onset, according to the widely known recommendations [8]. In our previous study, we did not use any cut-off value from ROC analysis, so we cannot provide any more comparisons. One may say that conformity at the level of around 80% is acceptable in daily practice but it is still low with regards to the ''any fracture'' criterion, which requires a special attention.

ROC analysis provides important data on the clinical utility of the used methods. In our previous study [16], we obtained AUC for any fracture at 0.833 and for FRAX US, 0.879 for Garvan nomogram, while for the hip, the respective values were 0.726 and 0.850, respectively. AUC values for women were significantly higher for Garvan nomogram. In the group of men, when compared with the group of women (see our previously published report), we obtained: lower AUC values for either method (FRAX or Garvan) in cases of any fracture risk assessment, while higher AUC values for FRAX and lower AUC values for the Garvan algorithm were noted in cases of hip fracture assessment risk. Generally, AUCs were lower for hip than for any fractures, just like in women. Our AUCs were higher than the respective values in a recent, 2-year prospective international study in 19 586 postmenopausal women [17]. However, in the cited study, BMD was not included.

Ten-year fracture probability was assessed in male populations by some authors [11–14]. In a study by Sandhu et al. [11], in a group of 56 men, FRAX US and FRAX UK were used along with Garvan algorithms. The general fracture risk was higher for Garvan nomogram than for FRAX algorithm results (no data were provided for hip fracture risk), what is comparable with our study. Compatibility with the used method was determined, using the correlation of 0.6 between FRAX and Garvan data. Unfortunately, no separate data were presented for hip and any fracture. Furthermore, no separate

data were presented for men and women and the concordance of fracture prediction was shown by simple correlation analysis, not as in our methodology, showing the conformity with regards to therapeutic decision thresholds. As expected, the Garvan model yielded a higher average prediction of major fracture occurrence in the fracture group, while the FRAX algorithm did not. In the reported study, we demonstrated that all the values of fracture prediction, established by the FRAX algorithm and Garvan nomogram, were significantly higher for all the variables. The authors also presented AUCs for Garvan and FRAX-US algorithms of 0.76 and 0.54, respectively. Unfortunately, no separate AUCs were provided for any or hip fractures. Irrespective of that, the AUC – as observed in our study – seemed to be higher. The authors drew a conclusion that the FRAX algorithm is a weak (rather useless) fracture risk assessment tool in men.

In a recent study, performed in 115 men, treated by androgen deprivation therapy (ADT) for localized prostate cancer, the necessity of treatment was verified, using low BMD and FRAX values [12]. The authors found that 33% of men on ADT had osteoporosis of spine, hip or forearm, confirmed by dual-energy X-ray absorptiometry (DXA), thus requiring an appropriate treatment. Using the FRAX tool in cases of corrected femoral neck, T-score identified only 17% of treatment demanding cases and, if calculated without femoral neck, 54% were identified as treatment needing cases.

In 363 men, treated with ADT for prostate cancer [13], a 10-year fracture probability was established by the FRAX algorithm, identifying a higher proportion of men in need of treatment than the traditional threshold of T-score −2.5 or less.

Recently, the performance of the FRAX algorithm system was independently assessed in a large clinical cohort of 36 730 women and 2873 men from the Manitoba Bone Density Program database [14]. In the 10-year Kaplan–Meier estimate for hip fractures in men, the observed risk was 3.5%, with predictive value of 2.9% and any fracture risk was 10.7% with predictive value of 8.4%. Fracture discrimination, based upon ROC curve analysis, was comparable to the published meta-analyses with the area under the ROC curve for osteoporotic fracture prediction of 0.694 (95% CI 0.684–0.705) for the FRAX algorithm with BMD and for hip fractures of 0.830 (95% CI 0.815–0.846), both of which were better than the FRAX algorithm results without BMD or with BMD alone. The authors concluded that the Canadian FRAX tool, calibrated on national hip fracture data, generated fracture risk predictions that were generally consistent with observed fracture rates across a wide range of risk categories. AUC values for any fracture were smaller than in our study (0.694 versus 0.808 and 0.843), while for hip fractures, AUC was higher (0.830 versus 0.748–0.749).

The reported study had several limitations. The Garvan nomogram values were calculated, based on the Australian male population, posing a certain risk of obvious differences between male populations in Australia and Poland. Because FRAX calculator produces estimates for ‘‘major fractures", that are limited to hip, humerus, spine and wrist, where the Garvan ‘‘all fracture’’ includes more fracture sites, the divergences between any fracture risks may, at least, be partially due to these methodological differences. In the FRAX algorithm approach, osteoporotic fracture(s) in adulthood should be taken into consideration, while in Garvan nomogram, fracture(s) after the age of 50 should be included. We did not verify fracture occurrence by the use of radiograms, thus certain fractures, especially silent spine fractures, could have been left unidentified. The authors of the FRAX algorithm propose to use it only in treatment-naive subjects, while a certain part of our population included men on antiresorptive therapy. They were included because we were not interested to follow-up patients and longitudinal modifications of fracture risk by the therapy do not interfere with a single comparison of both methods. We studied male outpatients only and this population may not be representative for the general population. One should also consider the fact that the lack of a cross-calibration procedure among the used densitometric scanners may have significantly affected the results of the study. And, finally, and important methodological difference between FRAX and Garvan methods should be presented. FRAX takes into account epidemiological data, including life expectancy in each population, while Garvan nomogram does not include this factor. Therefore, in fact, the FRAX algorithm establishes fracture probability (life expectancy is taken into consideration) and the Garvan algorithm estimated fracture risk (life expectancy is not taken into consideration). This important methodological difference probably decreases the conformity between both methods. This age-dependent discrepancy between methods may also be, at least partially, responsible for quite high ratio of men with low FRAX value despite of low T-score for femoral neck in the case of any fracture probability.

The part of the manuscript regarding the ROC analysis and comparing of AUCs for both methods would be perfect if we gathered longitudinal data in follow-up lasting for 10 years. Such longitudinal data would be necessary, if we aimed to establish sensitivity and specificity for each method separately. But, our aim in this part of the manuscript was to justify the thesis that each diagnostic tool needs to be validated individually and cut-off points (between high and low fracture risk) established for one method in one specific population cannot be directly implemented for other calculators or other ethnic groups. In such analysis, retrospective data are sufficient enough.

The question of a limited value of comparative studies, focusing on the validation of different fracture risk or probability calculators, was discussed in detail by Kanis et al. in the recently published opinion paper [18]. But, regardless of all the presented criticisms, one should agree that the direct comparison of different calculators is still justified by the fact that those diagnostic tools are dedicated to the same practical applications. The fact that they lead to different results makes a good starting point to understand the methodological differences in their design. It also provides a message that the estimations, obtained in each calculator, have to be interpreted in a bit different way, although they both are intended for the same application. This conclusion, derived from our study, seems to have a significant practical value.

Our study also has strengths: we were successful to get a relatively large study group of men in a wide age range, recruited from four medical centres. Also, the number of risk

factors, present in the studied men, enabled a reliable fracture risk and probability assessment.

Concluding, in general, both methods comparably predict fractures in regard to hip fracture. In the case of any fracture, the level of conformity between both methods is much lower (using the threshold of 20%), suggesting that the FRAX algorithm leads to fracture risk underestimation. However, when we reclassified our patients at low or high fracture risk level, according to the new cut-off points based on ROC analysis, the conformity for any fracture improved significantly. This may be suggestive of a somewhat misleading character of ''universal'' cut-off point, however, it is still necessary to establish a cut-off point, according to the results of the sensitivity-specificity analysis, separately for each method and for the reference population.

Declaration of interest

The authors declare they have no competing interests or other interests that might be perceived to influence the interpretation of the article.

References

1. Nguyen ND, Ahlborg HG, Centre RJ, et al. Residual lifetime risk of fractures in women and men. J Bone Miner Res 2007;22:781–8.
2. Kanis JA on behalf of the WHO Scientific Group. Assessment of osteoporosis at the primary health care level. Sheffield, UK: University of Sheffield; 2007.
3. Black DM, Steinbuch M, Palermo L, et al. An assessment tool for predicting fracture risk in postmenopausal women. Osteoporos Int 2001;12:519–28.
4. van Staa TP, Geusens P, Kanis JA, et al. A simple clinical score for estimating the long-term risk of fracture in postmenopausal women. QJM 2006;99:673–82.
5. Ettinger B, Hilier TA, Pressman A, et al. Simple computer model for calculating and reporting 5-year osteoporotic fracture risk in postmenopausal women. J Womens Health (Larchmt) 2005;14:159–71.
6. Hippisley-Cox J, Coupland C. Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores. Br Med J 2009;339:b4229.
7. Kanis JA, Johnell O, Oden A, et al. FRAXTM and the assessment of fracture probability in men and women from the UK. Osteoporos Int 2008;19:385–97.
8. Siris ES, Baim S, Nattiv A. Primary care use of FRAX: absolute fracture risk assessment in postmenopausal women and older men. Postgrad Med 2010;122:82–90.
9. Nguyen ND, Frost SA, Centre JR, et al. Development of a nomogram for individualizing hip fracture risk in men and women. Osteoporos Int 2007;18:1109–17.
10. Nguyen ND, Frost SA, Centre JR, et al. Development of prognostic nomograms for individualizing 5-year and 10-year fracture risks. Osteoporos Int 2008;19:1431–44.
11. Sandu SK, Nguyen ND, Centre JR, et al. Prognosis of fracture: evaluation of predictive accuracy of the FRAXTM algorithm and Garvan nomogram. Osteoporos Int 2010;21:863–71.
12. Adler RA, Hastings FW, Petkov VI. Treatment thresholds for osteoporosis in men on androgen deprivation therapy: T-score versus FRAX. Osteoporos Int 2010;21:647–53.
13. Saylor PJ, Kaufman DS, Michaelson MD, et al. Application of a fracture risk algorithm to men treated with androgen deprivation therapy for prostate cancer. J Urol 2010;183:2200–5.
14. Leslie WD, Lix LM, Johansson H, et al.; Manitoba Bone Density Program. (2010) Independent clinical validation of a Canadian FRAX((R)) tool: fracture prediction and model calibration. J Bone Miner Res 2010;25:2350–8.
15. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–45.
16. Pluskiewicz W, Adamczyk P, Franek E, et al. Ten-year probability of osteoporotic fracture in 2012 Polish women assessed by FRAX and nomogram by Nguyen et al.-Conformity between methods and their clinical utility. Bone 2010;46:1661–7.
17. Sambrook PN, Flahive J, Hooven FH, et al. Predicting fractures in and international cohort using risk factor algorithms without BMD. J Bone Miner Res 2011;26:2770–7.
18. Kanis JA, Oden A, Johansson H, McCloskey E. Pitfalls in the external validation of FRAX. Osteoporos Int 2012;23:423–31.