Author's response to reviews

Title: FRAX (R) tool, the WHO algorithm to predict osteoporotic fractures: an analysis of its discriminative and predictive ability in a Spanish female cohort (FRIDEX).

Authors:

Rafael AZAGRA (rafael.azagra@uab.cat)
Genis ROCA (groca@movistar.es)
Gloria ENCABO (gencabo@vhebron.net)
Amada AGUYE (amyaguye@telefonica.net)
Marta ZWART (martaifelix@hotmail.com)
Silvia GÜELL (24965sgp@comb.cat)
Nuria PUCHOL (39434npr@comb.cat)
Emili GENE (29833egt@comb.cat)
Enrique CASADO (guique.casado@gmail.com)
Pilar SANCHO (sanchomp@gmail.com)
Silvia SOLÁ (ssolamun@yahoo.es)
Pere TORÁN (ptoran.bnm.ics@gencat.cat)
Milagros IGLESIAS (milaiglesias@camfic.org)
M Carmen GISBERT (mariacarmengisbert@gencat.cat)
Francesc LÓPEZ-EXPÓSITO (flopeze@papps.org)
Yolanda FERNÁNDEZ (yolafern@yahoo.es)
Ana PUENTE (20892aps@comb.cat)
Mireia ROSÀS (mireiarosas@gmail.com)
Vicente BOU (vbou@gencat.cat)
Juan José ANTÓN (bilbojuanjo@hotmail.es)
Gustavo LANSDBERG (gustavolandsberg@gmail.com)
Juan Carlos MARTÍN-SÁNCHEZ (juancarlosmarsan@gmail.com)
Adolf DÍEZ-PÉREZ (ADiez@imas.imim.es)
Jesús PUJOL-SALUD (jpuiols@medynet.com)
Daniel PRIETO-ALHAMBRA (dprieto@idiapjgol.info)

Version: 11 Date: 31 August 2012

Author's response to reviews: see over
Barcelona, August 31th, 2012

Dear all,

Please find enclosed our manuscript “FRAX® tool, the WHO algorithm to predict osteoporotic fractures: an analysis of its discriminative and predictive ability in a Spanish female cohort (FRIDEX)”. This paper was already submitted at your journal and the current redaction includes the answer to Editor's comments.

We would appreciate your re-considering it for publication in the BMC Musculoskeletal Disorders as a Research Article. As requested in the Instructions for authors, we state that the manuscript reports an original primary research.

All the authors carefully read the manuscript and fully approve of it. The article is original. We would of course be ready to provide further information about our data and methods you so desire.

Correspondence about the manuscript should be addressed to me.

Thank you very much for your kind attention. We look forward to hearing from you.

Best regards,

Rafael Azagra

Family physician.
Badia del Vallès Health Centre, Catalan Health Institute
C/ Bética s/n,
08214 Badia del Vallès (Barcelona), Spain.
Tel: +34 937 192 600 Fax: +34 937 291 381
Email: rafael.azagra@uab.cat
Editor's comments:

I just would like to ask the authors change one thing. In the conclusion, they write that 'FRAX without BMD for major and hip fracture demonstrates a good discriminative capacity ...'. This is NOT true. The AUC value for FRAX without BMD was 0.693, and with this value, they can only claim 'poor' discrimination. Only when AUC values in the range of 0.8 and 0.9 they can claim 'good discrimination'. We would be grateful if you could address the comments in a revised manuscript and provide a cover letter giving a point-by-point response to the concerns.


Conclusions

FRAX™ without BMD for major and hip fracture demonstrates a good discriminative capacity with the AUC ROC for Spanish women but its predictive capacity does not adjust well with the current algorithm leading to underdiagnosis for major fracture and hip fractures. Simple models based on age or BMD alone predicted 10-year risk of major and hip osteoporotic fractures, as well as more complex FRAX™ models.

Conclusions

The current version of FRAX® for Spanish women without BMD analysed by the AUC ROC demonstrate a poor discriminative capacity to predict major fractures but a good discriminative capacity for hip fractures. Its predictive capacity does not adjust well because leading to underdiagnosis for both predictions major and hip fractures. Simple models based only on age or BMD alone similarly predicted that more complex FRAX models.

Main text. Discussion, page 17. Third paragraph.

The FRAX tool can therefore be considered to present with a poor discriminatory capacity for women to have major osteoporotic fractures within 10 years, with this capacity being good for hip fractures without the need of determining the BMD, although this improves somewhat with its determination. The FRAX tool shows a scarce predictive capacity of the risk of fracture and predicts less than 50 % of those which occur. The reason for this underdiagnosis may be because the Spanish cohort introduced as the
reference in the FRAX tool is not representative of the current female population since these women present significantly more fractures than those actually predicted by the FRAX tool.

**Main text. Discussion, page 19. First paragraph.**

In summary, as a conclusion, FRAX without BMD demonstrates a poor discriminative capacity for major fractures and a good discriminative capacity for hip fractures with the AUC ROC for Spanish women but its predictive capacity does not adjust well with the current algorithm leading to underdiagnosis for major fracture and hip fractures. On introducing the values of the L1-L4 T-score in the FRAX tool, the result did not provide an improvement in the discrimination of vertebral fractures measured with the AUC-ROC. Simple models based on age or BMD alone predicted 10-year risk of major and hip osteoporotic fractures, as well as more complex FRAX models.
Reviewer (1) ’s report (R1)’

Comments on “FRAX tool, the WHO algorithm to predict osteoporosis fractures: an analysis of its discriminative and predictive ability in a Spanish cohort (FRIDEX)

Major Compulsory Revisions

1. Please provide information on how the following study variables were defined and/or measured: height, weight, body mass index, smoking status, alcohol risk, previous fracture, parental osteoporosis or fractures, glucocorticoids, rheumatoid arthritis, calcium/vitamin D supplements, active bone drugs. For example, it is unclear how alcohol risk is defined (i.e., how many drinks or what amount of alcohol consumption is used to define risk?). If all of the measures were collected by self-report questionnaire, then the authors needs to discuss the potential limitations of this approach. For example, socially desirable responses may result in substantial under-reporting of current smoking status. Appropriate references should be cited for the variable definitions and the questionnaire design.

Updated.

Height, weight, body mass index were obtained during baseline DXA scan. The rest of baseline items were obtained by semi structured questionnaire by interviewer during the same visit. On the other hand, the variables are set according to the instructions of the official website of FRAX (http://www.shef.ac.uk/FRAX/tool.jsp?lang=sp). The variables which are mentioned in the questionnaire were defined as well according to standard units of measurement for each. Regarding the risk of alcohol consumption, the quantification of consumption in standard drinks (UBEs) allows rapid quantification of consumption and its easy conversion into grams of pure alcohol. The value of the UBE in Spain with a slight North-South gap is set to 10 g of alcohol and is equivalent to a consumption of wine (100ml), sparkling wine (100ml) or beer (200 ml) half and consumption of distilled or combined (25 ml). Weekly risky drinking for women and over 65 years is that is> 17 UBEs and men> 28- UBEs. The phone records of alcohol consumption have shown good validity and correlation in Mediterranean countries where alcohol consumption is widespread. [page 8].

Regarding the consumption of snuff, not only made reference to the number of cigarettes smoked per day, but also the time spent as a smoker, which focuses on some characteristics of tobacco use. These two data are included in the concept: number packages/year. To determine this figure performs a simple operation: multiply the number of cigarettes smoked per day by the pool by the number of years been consumed that amount of snuff and the result is divided by 20.

Reference:

2. For what proportion of individuals was a proxy response obtained from a
relative or care provider? Updated.
Only in case of personal circumstances (deafness, slurred speech, etc.) a part
of the information was obtained through regular cohabiting relatives of patients
in 15 of 770 cases (1.9%). [page 8]

3• The authors used the telephone questionnaire to obtained information on
fragility fractures and then medical records were reviewed to confirm the validity
of the fracture self-report. Please discuss the sensitivity and specificity of the
fracture self-report when compared to the medical record. The use of self-report
could result in under-reporting of fractures, particularly for clinical spine
fractures; please discuss this as a potential limitation of the study.
Updated.
Self-reported generally have a significant correlation with those in the medical
record. In any case always been found documented as explained. In all cases
of fracture the medical records of the patients were reviewed and, when
necessary, we requested a medical report for its validation. All cases of fracture
that could not be verified or those arising from a motor vehicle accident or major
trauma were excluded from analysis, fractures in the history of the subjects
under study. A potential limitation of self-reported fractures is in vertebral
fractures. In our study the total self-reported fractures were 16% higher than
they were registered and so were excluded from the final analysis. It can be an
advantage for risk predictions proposed by FRAX™. [page 12]

4• Table 1 compares the baseline characteristics of the study non-participants
with the study participants before several exclusions were made. It is more
appropriate to include participants after these exclusions; Table 1 should be
revised accordingly.
Updated.

5• On page 12, the authors note that “The lower part of the figure [Figure 4]
represents the same results after multiplication (calibration) by the number of
times that the ObsFx is greater than the ExpFx (table 5).” The authors should
describe the simulation methodology in the analysis plan.
Updated.
The proportion of fractures expected is calculated by the sum of an individual
probability of fracture from all women included/100. Model calibration is done by
multiplying the FRAX values by the ratio ObsFx/EspFx. [page 9]

6• On page 18, the authors note that FRAX tool is based on DXA values of the
femoral neck and that prediction may change with the use of the lumbar spine
T-score. One relevant reference for discussion is found in Osteoporos Int; 2011;
22:839–847; Spine–hip discordance and fracture risk assessment: a physician-
friendly FRAX enhancement by W. D. Leslie et al.
Updated.
Although it has described that a correction can adapt the lumbar spine BMD
and improve the prediction for major and vertebral fractures of FRAX (Leslie
WD, 2011), in our study by incorporating the lumbar spine BMD did not improve
the discriminative ability of FRAX measured by AUC with femoral neck BMD neither for major or vertebral fracture (data not showed).

7. The authors have not described how the relative risk values in Tables 2 and 3 were computed. This information should be included in the analysis plan. Updated. The RR was calculated by quotient between prevalence of each risk factor in fractured women and in non-fractured. [page 9]

Minor Essential Revisions

8. Figure 1: the centre set of numbers in the flow chart should be accompanied by descriptive labels; the figures for male and female at the bottom of the page should be replaced with written labels. Updated. New figure 1.

9. Figures 2 and 3: the numbers reported above the ROC curves (i.e., at the top of the page) should be accompanied by descriptive labels: AUC ROC; 95% confidence interval; p-value. However, Table 5 repeats most of the information already reported in Figures 2 and 3. Therefore, the authors should carefully consider the value of including these figures. Updated. We removed the values are not described in the figure and keep the table 5

10. Figures 4 and 5: the word “linear” is mis-spelled in the two right-hand panels; the text within the graphs is in a font that is too small to be easily read and therefore the font size should be increased. Updated.

11. Figures 4 and 5: please provide figure notes that explain what is represented by the dashed line and the solid lines. Updated.

12. Table 1: the sample size values should be accompanied by the labels “n =”. Updated.

13. Table 2: The headings for columns 2 and 3 should be revised: “Women with fracture (n = 65)” and “women without fracture (n = 705)”. The second and third columns in Tables 2 and 3 are the same. The two tables can therefore be combined in order to reduce the total number of tables that accompany the manuscript. Updated. Modified

14. Abbreviations are not used consistently throughout the manuscript. For example, on page 7, the abbreviation TQ is defined, but then this abbreviation is not used in place of “telephone questionnaire” on page 10 in the first paragraph (5th line). Updated.
Discretionary Revisions None

Reviewer (2) ’s report (R2)

Version: 4 Date: 20 February 2012
Reviewer: Kim Brixen
Reviewer's report:

The paper reports a prospective, observational study with 10 years follow-up comprising a random sample of 770 women aged 40-90 referred for DXA.

Baseline data were acquired with DXA and by a questionnaire. Follow-up data on incident fractures were collected by telephone interviews validated against case records. A total of 82 fractures in 65 women were recorded including 18 hip fractures in 17 women. The main conclusion is that FRAX discriminates well, however, the Spanish edition needs to be adjusted to be accurate.

Aim. The aim of the study is clearly stated and this represents a valid research question based on previously published studies in the literature. Given the current interest in FRAX and need for independent validation of FRAX, the issue of high interest.

Design. The study design (prospective, observational) is the best to address the hypothesis.

Participants. Details on the inclusion of participants are not completely clear and the particular cohort was not ideal to address the question at hand. The total FRIDEX-cohort comprised 34,636 persons of both genders. However, the interesting figure would be the number of FRIDEX-participants potentially available with 10 years of follow-up.

We have removed these data about all persons in the cohort because they are not relevant to this study. As suggested by the reviewer is important to describe women who are 10 years in the cohort. Answer: Page 8; Paragraph 2: The cohort available over 10 year follow-up was 3459 people: 142 men and 3437 women, as shown in figure 1.

Follow-up and outcome. The follow-up period (10 years) was adequate. However, the acquisition of data regarding fractures was less than perfect since this was based on telephone interviews and no data could be retrieved in 36% of the cases and total of 5% of the initial cohort died during follow-up. Fractures were, however, verified against patient records.

Confounders and bias. Fracture data were captured in retrospect. Potentially, this introduces a recall-bias. Participants were selected at random between
patients referred for DXA on the basis of risk factors. This introduces a bias – the cohort a priori has a higher risk of fracture and the prevalence of risk factors in the cohort must be higher (and possible skewed) as compared with the general population. This is to some extend discussed by the authors.

Statistical analyses. ROC analysis was used to evaluate sensitivity and specificity of FRAX. This approach is problematic although many studies have done the same.

FRAX predicts a risk between 0 and 1. A particular patient may thus have a risk of e.g. 0.3 for suffering a fracture during the next 10 years. ROC, however, depends on dichotomies and it could only be used correctly if FRAX predicted that an individual patient to have the risk of 0 or 1 for fracture. In other words, for an individual patient the predicted risk may be of any value between 0 and 1 while the observed risk fracture in the same patient is either exactly 0 or exactly 1 after 10 years. So was the predicted risk of 0.3 wrong if the patient did not fracture?

Performance of FRAX may be evaluated in groups of patients as done by the authors in the section on the Hosmer-Lemeshow approach. Obvious questions are not addressed; is the predictive power of FRAX superior to that of age and sex alone (see below)?

Discussion. The discussion is kept to the point; however, a few issues are neglected (see below).

Conclusion. Apparently, no conclusion is stated at the end of the discussion, albeit a conclusion – drawn within the boundaries of the presented data - is made in the abstract.

Ethics. Ethical standards were observed.

Style. The paper is well written in a short a concise and straightforward style.

The authors, however, do not explicitly follow the guidelines for reporting observational studies (e.g. STOBE).

This work follows the guidelines of the STROBE initiative for cohort studies.[page 1, 9]

Comments are welcome.

Reviewer (2)

Major issues

1. The authors should discuss the inherent problems with ROC analyses in the present setting.
The ROC curve has several problems. For analysis of sensitivity and specificity we have not a gold standard of FRAX™ for Spanish population. Moreover, ROC needs a gold standard of illness (fracture) and we do not have because of the electronic records are not completely reliable and we needed to make a double check (self-reported validate against records). On the other hand, the area under the ROC curve is important, since it measures the discrimination power of the model. Nevertheless, tests of discrimination alone are not sufficient for model evaluation, since they do not indicate whether calibration is also good.

2. How were data on mortality collected? It seem that no fracture data were available in those patients who died during follow-up?

Dead patients were not studied because of the impossibility of obtain all the study variables and to answer the questionnaire by relatives. [page 7]. About 69 patients died, 43.4% (39) have been detected by searching the telephone number and detect the death. In the other 30 cases were detected through contact with family and reported only 2 cases of fracture between baseline and the date of death. [page 10]. Other potential confounders and biases are that we excluded those who died during the follow-up [page 17]

3. Fracture data was captured in retrospect. This introduces the possibility of recall-bias. This issue should be discussed.

Other potential confounders and biases were that the collection of incident fractures is captured in retrospect, the validation records was only for patients with fractures and, as well, usually the electronic registers of fracture tends to be less records than actually occur. To minimize these potential biases we have verified all self-reported fractures and not included in the study which did not fulfill both (self-reported and recorded). Therefore, this type of analysis tends to benefit the predicted fractures in the ratio ObFx/ExpFx. [Page 16-17]

4. A total of 18% of the participants received vitamin-D and calcium supplementation. How did this affect the fracture risk and evaluation of FRAX? The potential confounding should be discussed.

Other potential confounding factor can be the Calcium/Vit D supplement intake because we have not excluded at baseline or during the study period. There is important discussion in the literature about the role of these supplements in reducing the risk of fracture, except in a subgroup of patients taking bone active drugs for the potential hypocalcaemia or in patients admitted to nursing homes. These patients are not included in this study. Moreover there is no significant difference between Calcium/Vit D supplement intake between participants and no participants. [Page 15]

5. A total of 29% of the potential participants were excluded due to treatment with specific anti-osteoporotic medication. Presumably, these were the patients with the highest risk of fracture. How did exclusion of such a large proportion of
participants affect the evaluation of FRAX? The potential confounding should be discussed.

FRAX has been proposed by its promoters as a tool to determine the risk in patients without active bone drugs. However, we introduce as potential confounders in the discussion with the following sentence: We have excluded from the analysis of the cohort of women receiving active treatment for the bone at baseline of the study because the FRAX tool has so defined, but we have not been excluded women who received treatment during the 10-year period. This can be a potential confounding factor, however exclude women would mean removing the greatest potential for fracture, but keep going who have received treatment can be reduced the all risk of new fractures observed.[page 14-15]

6. It is strongly suggested to expand the analysis to include the performance of e.g. age and sex. Previous studies have suggested that age performs virtually as well as FRAX. The present dataset could be of interest to validate this finding. Such analyses could alter the conclusion significantly.

We compare AUC of the ROC curve of FRAX for major and hip fracture with a simple models including only age. The AUC in a model that includes only age was 0.668 for major fracture and 0.882 for hip fracture with no significant differences with the results of FRAX (p=0.565 and p=0.976 respectively). [page 11]. Simple models based on age or BMD alone predicted 10-year risk of major and hip osteoporotic fractures, as well as more complex FRAX models.[conclusions, page 1-2 and 16]

7. The potential selection bias that could result in the cohort having a higher risk of fracture and a higher prevalence of risk factors than the general population is addressed by the authors in the discussion, however, the sentence “However, there is no evidence that patients pre-selected strategically may have a greater risk than the general population” is inappropriate – the burden of proof is on the authors.

We have removed this sentence. “However, there is no evidence that patients pre-selected strategically may have a greater risk than the general population.” We have added the following. However it is important to know the profile of women who are selected to perform the DXA-scan by general practitioners and other specialists as may higher but close to the general population over 50 years. [Page 15-16].

8. A conclusion should be stated at the end of the discussion (as in the abstract).

In summary, as a conclusion, FRAX without BMD for major and hip fracture demonstrates a good discriminative capacity with the AUC ROC for Spanish women but its predictive capacity does not adjust well with the current algorithm leading to underdiagnosis for major fracture and hip fractures. In this study, on introducing the values of the L1-L4 T-score in the FRAX tool, the result did not provide an improvement in the discrimination of vertebral fractures measured with the AUC-ROC. We advise our Spanish colleagues to use the FRAX tool in clinical practice but weighing the resulting value of each individual case of the FRAX tool without BMD by a calibration value to obtain
an absolute risk value of major or hip fracture at 10 years. New studies may allow a single value of calibration which is easier to remember in clinical practice. The result obtained will be more adjusted to the reality of the risk of fragility fracture in our country according to the results found in the present and other studies. [Page: 16].

9. Adherence to guidelines such as STROBE is strongly advised to improve published papers in general. The present paper could easily be adapted to such standard. This work follows the guidelines of the STROBE initiative for epidemiological studies [http://www.strobe-statement.org/index.php?id=strobe-publications].[page 1,9].

Minor issues

10. Subheadings in the introductions should be removed.
Updated.

11. It is suggested that a note is made on the fact, that the FRAX algorithm is neither published nor freely available for independent researchers.
Updated at page 6, 3rd paragraph.

12. In the abstract (page 3), the number and age distribution of participants, duration of follow-up should be stated. Similarly, it should be stated in the abstract how fractures were ascertained.
The study subjects were 770 women from 40 to 90 years of age in the FRIDEX cohort. The mean age was 56.8 ± 8 years. The fractures were determined by telephone questionnaire and subsequent testing in medical records at 10 years.

13. On page 5, the sentence "For many years the clinical, social and economic importance of osteoporotic fractures has been known to favour the incidence of new fractures and lead to disability [2]" is confusing. As is stands, the sentence implies that social and economic factors may be important risk factors for fracture. The quoted paper, however, deals with the consequences of fractures. The sentence should be rephrased or supported by other data.
Updated. It is well known that osteoporotic fractures involve a higher incidence of new fractures and lead to disability [2].

Updated.

15. On page 4, the statement “This has been demonstrated in different international studies” should be supported by reference to such studies.
Updated. We removed the following part of the text: “This has been demonstrated in different international studies”.

16. On page 5, the sentence “The objective of this study is to evaluate..” should read “The objective of this study was to evaluate..”. Updated.

17. On page 5, the sentence “The protocol, procedures and main characteristics of the study has recently been published..” should read “The protocol, procedures and main characteristics of the study have recently been published..” Updated.

18. On page 10, the sentence “except that the participants are a mean of one year younger” should read “except that the participants were one year younger on average”. The sentence “table 2 described the main characteristics” should read “table 2 describes the main characteristics”. Updated.

19. It should be stated clearly in the methods section, that the Spanish edition of FRAX that was applied. The FRAX™ tool was determined through the official website (version 3.2 accessed on October 2010). [page 8]

20. Abbreviations should be used consistently once defined. On page 12 “That is, of Bone Mineral Density by DXA” should read “That is, of BMD by DXA”. Similar adjustments should be made throughout. Updated.

21. On page 14, the sentence “… that they were one year older and were taking a larger percentage of glucocorticoids..” should read “… that they were one year older and more patients were on glucocorticoids..” Updated.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests: No competing interests