PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Assessing fracture risk in people with MS: a service development study comparing three fracture risk scoring systems
AUTHORS	Dobson, Ruth; Leddy, Sara; Gangadharan, Sunay; Giovannoni, Gavin

VERSION 1 - REVIEW

REVIEWER	Elisabeth Gulowsen Celius, MD,PhD Dep of Neurology Oslo University Hospital, Oslo Norway I have received funding for travel and speaker honoraria from Sanofi-Aventis, Merck Serono, Biogen Idec, Teva, and Novartis; and receives research support from Biogen Idec and Novartis.
REVIEW RETURNED	16-Jan-2013

GENERAL COMMENTS	This is a very interesting study adressing an important issue. There are although some minor concerns. Regarding the patient population Table 2 should include both EDSS and disease duration. Also the proportion of patients using walking aids should be given, were any patients wheel-chair-bound? In the text it seems that 49% of the patients have a history of falls, but in Table 2 the figure is 43.2%? This inconsistency should be checked. Also in the description og the subjects in the results section the figures 42/100 and 49/100 is given - the results for the 88 included would be important to report. The problem with the age limit of the FRAX and QFracture scores is adressed. As only half of the patients were below 40 this is a major concern and it would be interesting if the calculations could be performed only for the patients above the actual age limits. The conclusion of this study is reasonable, but it would be interesting. Also the influence of disease modifying drugs on fracture risk and these scores should be adressed in the discussion as well as the number of patients receiving DMDs (and duration) should be given. The start of the discussion seems a bit odd? A lost sentence?

REVIEWER	Eugene McCloskey Professor of Adult Bone Diseases Academic Unit of Bone Metabolism University of Sheffield. CoI - I am one of the developers of FRAX
REVIEW RETURNED	21-Jan-2013

THE STUDY	The major flaw in this paper is the same as that ignored by NICE in the recent Short Clinical Guideline. The calibration of the 3 tools differ (for example, QFracture is calibrated to fracture incidence recorded in the GP databases - an underestimate - while FRAX is calibrated to fracture incidence in the general UK population). For this reason, the results of the three tools will always have systematic differences between them so that the use of arbitrary thresholds (e.g. 5% for BMD) is inappropriate.
RESULTS & CONCLUSIONS	It would be better to examine ranking of the patients by each of the tools arther than comparing the absolute risk values for the reasons given above.
GENERAL COMMENTS	I think this is a useful process but needs to take account of design and calibration differences in the tools that will always give differences between them. The message that the results of all of these tools cannot be directly compared, or one substituted for another, is a really very important one. Each tool will need its own thresholds developed (as with the NOGG guideleine for FRAX)

REVIEWER	Niamh Cummins, PhD Research Fellow Graduate Entry Medical School University of Limerick Ireland No competing interests
REVIEW RETURNED	22-Jan-2013

THE STUDY	The following limitations should also be included; • The sample size of n=88 is relatively small and therefore should be acknowledged as a limitation of the study • The differing age restrictions of the risk calculators is also a limitation as it does not allow for direct comparisons. For example the youngest patient in the study was aged 22y and would have been entered as such in the MS-specific calculator however the same patient would have been entered as aged 30y in the QFracture calculator and age 40y in FRAX (the respective lower age limits). This also needs further clarification in the Methodology section (pg 7 lines 7-14). While fracture risk would not decrease significantly between the ages of 20 and 40y this is still a limitation of the analysis and should be acknowledged as such.
RESULTS & CONCLUSIONS	The results of the study are in agreement with previous observations in relation to comparisons between the risk calculators. FRAX has previously been compared to QFracture in a number of studies (although not in an MS population) and FRAX has also been compared to the MS-Specific score by Bazelier et al 2012. Reference should be made in the discussion to some of these previous studies.
GENERAL COMMENTS	This manuscript sets out to compare fracture risk screening tools in multiple sclerosis (MS) patients. This has not previously been well described in the MS population and therefore the authors attempt to investigate it further is a worthy endeavour. The manuscript is well written and the findings are interesting. However in addition to my previous recommendations in relation to the study limitations and discussion the following revisions should also be applied; Methods • In the statistical analysis section (pg 7, line 54) a reference is made

	to ELISAs. This is slightly worrying as it may suggest copying and pasting from a previous paper, however the statistical tests used do appear to be appropriate. The typo should be corrected. Results There are some errors in relation to naming of the tables and figures in the Results section • Pg 8 lines 54/55 figure 2 should read figure 2a • Pg 9 line 3 figure 3 should read figure 2b • Pg 9 lines 10/11 table 4 should read table 3 • Pg 9 line 42 figure 4 should read figure 3

VERSION 1 – AUTHOR RESPONSE

Reviewer 1:

Reviewer: Elisabeth Gulowsen Celius, MD,PhD Dep of Neurology Oslo University Hospital, Oslo Norway

I have received funding for travel and speaker honoraria from Sanofi-Aventis, Merck Serono, Biogen Idec, Teva, and Novartis; and receive research support from Biogen Idec and Novartis.

This is a very interesting study addressing an important issue.
There are although some minor concerns. Regarding the patient population Table 2 should include both EDSS and disease duration.

Unfortunately full EDSS data on fully ambulatory patients was not available as part of this study. Data regarding disease duration has been added (page 8 paragraph 2 and table 2). Data regarding disease modifying treatment has also been added (page 8 paragraph 2 and table 2).

Also the proportion of patients using walking aids should be given, were any patients wheel-chair-bound?

4 patients reported using a wheelchair at any time but were ambulatory for >20m with bilateral assistance (EDSS 6.5), only one patient used a wheelchair all of the time (i.e. EDSS 7.5). This has now been made clear in the results section (page 8 paragraph 3) and the data added to table 2.

In the text it seems that 49% of the patients have a history of falls, but in Table 2 the figure is 43.2%? This inconsistency should be checked.

This inconsistency has been checked and was due to a typographical error. Rather than 38/88 patients falling, the correct figure was 48/88 (i.e. 54.5%). The 49/100 figure in the text refers to all 100 patients rather than just the 88 on whom fracture risk data was available. This has hopefully now been made clearer in the text (page 8 paragraph 3), with figures provided for both the full cohort of 100 and also the 88 on whom the fracture risk data was calculated.

Also in the description of the subjects in the results section the figures 42/100 and 49/100 is given - the results for the 88 included would be important to report.

This important information has now been added (page 8 paragraph 3).

The problem with the age limit of the FRAX and QFracture scores is addressed. As only half of the patients were below 40 this is a major concern and it would be interesting if the calculations could be

performed only for the patients above the actual age limits. The conclusion of this study is reasonable, but it would be interesting.

This analysis has been performed, using the 38 patients aged 40 or over. Despite selecting only the patients who are within age range for all tests, the highly significant difference between the three tests remains ($p<0.0001$ for all comparisons). This is true for both the overall fracture risk estimate and the hip fracture estimates. This has been discussed in the methods (page 7 paragraph 1), and the results given (page 9 paragraph 3 and page 10 paragraph 3). A section has also been added into the discussion to mention the implication of these results (page 11 paragraph 3). Figures 1b and 3b have been added to illustrate this.

Also the influence of disease modifying drugs on fracture risk and these scores should be addressed in the discussion as well as the number of patients receiving DMDs (and duration) should be given.

Details regarding disease modifying agents have been added to table 2. A section regarding the influence of disease modifying drugs has been added to the discussion (page 13 paragraph 1).

The start of the discussion seems a bit odd? A lost sentence?

The start of the discussion has been reworded.

Reviewer 2:

Reviewer: Eugene McCloskey
Professor of Adult Bone Diseases
Academic Unit of Bone Metabolism
University of Sheffield.

CoI - I am one of the developers of FRAX

The major flaw in this paper is the same as that ignored by NICE in the recent Short Clinical Guideline. The calibration of the 3 tools differ (for example, QFracture is calibrated to fracture incidence recorded in the GP databases - an underestimate - while FRAX is calibrated to fracture incidence in the general UK population). For this reason, the results of the three tools will always have systematic differences between them so that the use of arbitrary thresholds (e.g. 5% for BMD) is inappropriate.

See comments above. It would be better to examine ranking of the patients by each of the tools rather than comparing the absolute risk values for the reasons given above.

This is an extremely useful comment. The fact that it has not been taken into account by NICE means that many clinicians may also not be aware of this. It has been added to the discussion section, and the authors are grateful to the reviewer for bringing it to their attention.

Using the dataset for this population, rank quintiles for each fracture risk calculator have been generated. This is an accepted method to examine ranking that has been previously used in the MS literature (Levin et al 2005, Munger et al 2006). Using the rank quintile allows the three scores to be compared in pairs regardless of calibration. This has been done, and the methods discussed (page 8 paragraph 2). The results of this comparison are given (page 9 paragraph 4 and page 10 paragraph 3).

I think this is a useful process but needs to take account of design and calibration differences in the tools that will always give differences between them. The message that the results of all of these tools cannot be directly compared, or one substituted for another, is a really very important one. Each tool will need its own thresholds developed (as with the NOGG guideline for FRAX).

Again, this is an extremely useful and important point. It has been added to the discussion (page 11 paragraph 2 and page 12 paragraph 3- page 13).

Reviewer 3:

Reviewer: Niamh Cummins, PhD

Research Fellow
Graduate Entry Medical School
University of Limerick
Ireland

No competing interests

The following limitations should also be included; • The sample size of n=88 is relatively small and therefore should be acknowledged as a limitation of the study •

This has been acknowledged in the discussion (page 11 paragraph 3)

The differing age restrictions of the risk calculators is also a limitation as it does not allow for direct comparisons. For example the youngest patient in the study was aged 22y and would have been entered as such in the MS-specific calculator however the same patient would have been entered as aged 30y in the QFracture calculator and age 40y in FRAX (the respective lower age limits). This also needs further clarification in the Methodology section (pg 7 lines 7-14). While fracture risk would not decrease significantly between the ages of 20 and 40y this is still a limitation of the analysis and should be acknowledged as such.

I agree regarding this is a major limitation inherent in the calculators. The fact that patients aged <40 would have had different ages assigned in at least two of the calculations has now been made clearer in the methods (page 7 paragraph 1). Given this limitation, and the comments by reviewer #1, a further subgroup analysis has been performed including only those patients aged 40 or over. Whilst this analysis is limited by a small sample size of 38 (the small sample size is made clear page 9 paragraph 3 and page 11 paragraph 3), the discrepancies between the three fracture risk calculators remain for 10 year risk of both any fracture and hip fracture.

The results of the study are in agreement with previous observations in relation to comparisons between the risk calculators. FRAX has previously been compared to QFracture in a number of studies (although not in an MS population) and FRAX has also been compared to the MS-Specific score by Bazelier et al 2012. Reference should be made in the discussion to some of these previous studies.

References have been added to illustrate the previous attempts to compare FRAX and QFracture, together with a reference to a review article discussing this in more detail (page 11 paragraph 2). The MS-specific fracture risk score was compared to FRAX by Bazelier et al in the paper describing this score, and this has been commented on in the discussion (page 11 paragraph 2). However, we could

not find any additional pubmed or scopus listed articles comparing the MS-specific score to either FRAX or QFracture. Should the reviewer be aware of any publications that we have missed then we would of course be happy to include these.

This manuscript sets out to compare fracture risk screening tools in multiple sclerosis (MS) patients. This has not previously been well described in the MS population and therefore the authors attempt to investigate it further is a worthy endeavour. The manuscript is well written and the findings are interesting. However in addition to my previous recommendations in relation to the study limitations and discussion the following revisions should also be applied;

Methods
• In the statistical analysis section (pg 7, line 54) a reference is made to ELISAs. This is slightly worrying as it may suggest copying and pasting from a previous paper, however the statistical tests used do appear to be appropriate. The typo should be corrected.

Apologies for this typographical error, this has been corrected.

Results
There are some errors in relation to naming of the tables and figures in the Results section • Pg 8 lines 54/55 figure 2 should read figure 2a • Pg 9 line 3 figure 3 should read figure 2b • Pg 9 lines 10/11 table 4 should read table 3 • Pg 9 line 42 figure 4 should read figure 3

Again, apologies for these typographical errors, these have now been corrected.

VERSION 2 – REVIEW

REVIEWER	Elisabeth Gulowsen Celius, MD, PhD Dep of Neurology Oslo Universtiy Hospital, Ullevål 0407 Oslo Norway
REVIEW RETURNED	13-Feb-2013

GENERAL COMMENTS	Questions from previous review are all thoroughly answered, recommends acceptance.