Author’s response to reviews

Title: The effect of variation in interpretation of the La Trobe radiographic foot atlas on the prevalence of foot osteoarthritis in older women: the Chingford general population cohort

Authors:

Peter McQueen (plm1g09@soton.ac.uk)
Lucy Gates (L.Gates@soton.ac.uk)
Michelle Marshall (m.marshall@keele.ac.uk)
Michael Doherty (Michael.Doherty@nottingham.ac.uk)
Nigel Arden (Nigel.Arden@ndoms.ox.ac.uk)
Catherine Bowen (c.j.bowen@southampton.ac.uk)

Version: 1 Date: 31 Oct 2017

Author’s response to reviews:

Reviewer #1: General comment:

The study tests if there are differences in radiographic OA prevalence for different scoring approaches of the LFA. The study concludes that there needs to be harmonisation of case definitions of foot OA. The case definition (ie: a score of 2 for OP or JSN in any view) has not been varied. Therefore, can the authors please consider revision of this statement - there should be harmonisation of scoring approaches… Please review the Abstract, Introduction and Discussion/conclusion regarding this wording.

Response: We would like to thank the reviewer for their thorough and thoughtful comments towards improving the clarity of our manuscript.

We have considered all comments and revised the wording throughout.

My comments are below

Abstract:

1. As LFA is stated in the Background, it can be used within the Methods rather than 'La Trobe Foot Atlas'. There appears to be inconsistency in use of hyphen versus en dash / em dash.
Response: This has been corrected

Background:

2. P4L77: Please consider adding of hand OA figures given the sentence is a generic one as to the disability caused by OA?

Response: We have altered the first paragraph to include the hand. “The prevalence of radiographic hand OA has been reported to range from 27% to 83% (Pereira et al; Larence et al). Hand OA is said to consist of several phenotypes that make it more complex to study (Zhang et al). Whilst investigation of foot joints maybe more aligned to those of the hand, the prevalence of radiographic and symptomatic radiographic OA is much less understood”.

3. P4L83: Please include some description of the ages associated with the prevalence of symptomatic foot OA. Could this sentence be moved to the concluding sentence of para 1?

Response: This information has been added and we have moved the sentence.

4. P4L86: Do references 12, 13 and 14 support your argument regarding discordance in how foot OA is defined?

Response: Thank you for pointing this out. We have removed the word foot, as we agree the emphasis is on discordance in defining OA of the knee and hip. We have added “For investigations of foot OA,” at the start of the following sentence.

5. P4L93: Please check for missing full stop.

Response: This has been amended.

6. P4L101: Please review rogue space ‘( 1st MTPJ)’.

Response: This has been amended.
7. P5L108: Can you please provide some references and explanation to support your notion that '…as with other radiographic atlases, the ordinal technique for scoring introduces an interpretative approach….under or over estimation in the prevalence of OA'.

Response: We have amended the paragraph for clarity to:

“The presentation of radiographic features varies quite widely. As radiographic atlases use semi-quantitative or ordinal grading systems to classify individuals, often into 4 or 5 categories, a degree of interpretation is required in order to categorise OA features”

“This is particularly likely when an unclear view of a joint is being assessed, which happens often for views of the midfoot and certain hind-foot joints. The authors of the original atlas themselves do suggest from their inter-rater reliability results that “there is some degree of inherent variability in the interpretation of some aspects of the atlas” Menz et al. [2007] and this has also been noted with other radiographic atlases [Kohn 2016]. We wished to evaluate how much this variation in interpretation can affect the prevalence of radiographic foot OA”.

8. P5L108: It may be worth including that the LFA uses an ordinal scoring system for JSN and OPs in the preceding paragraph where the LFA is first introduced.

Response: Thank you, we agree and have added that to the previous paragraph.

Methods:

9. P5L124: please add ' years' to after '45-64'.

Response: This has been amended.

10. Foot radiographs:

Can you please provide some data regarding the mean (SD) age, height, weight, BMI of your sample (218 women for which you had radiographs). This is important as the prevalence of foot OA does vary according to age (and BMI).

Response: Yes, we have added the mean (SD) age: 75.5 years (5.1); weight: 69.2kg (12.6); height: 158.4cm (6.1); BMI: 27.6 (44.75).
Radiographic scoring:

11. Please be more explicit who did the scoring of the radiographs.

Response: We have added the sentence: “All radiographs were scored by a single researcher (PMc)”.

12. P6L153: Please consider changing the word 'case definition' to 'scale description' (rather than case definition which relates to the presence of absence of OA at the joint).

Response: This has been amended.

13. P7L165: Please consider a subheading. Can you please provide some justification for the development of Technique 2. Has this approach been used previously? Technique 3 appears to be the same as Technique 1. Since, in the original development of the LFA atlas, the assessors gave a score that was 'conservative' where they were unsure (ie: a score sat in the middle of two scoring options).

Response: We have added the subheading ‘scoring techniques’.

Yes, technique 2 was a new approach not used previously. We have included this descriptor within that section and clarified as follows:

“It was devised by our team of experts in the field of osteoarthritis (NKA, MD), radiography (MM) and foot and ankle research (CB) to understand how the prevalence estimate of foot OA changed when a scorer did not include an estimate for any joint they could not make a decision on”.

We thank the reviewer for pointing out our typographical error in the explanation of technique three. This now reads as follows:

“Technique 3 was a revised version of Technique 1 whereby all joints that were difficult to interpret and score for OPs and/or JSN were given a score based on an over-estimate. (For example, where an OP in a participant’s joint may have been ambiguous to score between a grading of 2 or 3, the higher grading of ‘3’ was accepted)”.

14. P7L177: Please consider some type of 'reproducibility of radiographic scoring of foot osteoarthritis' subheading. Can you provide some inter-rater reproducibility data between PMc and the OA scorer who had previously used the LFA? Please clarify who was this person. The reproducibility of the rater (PMc) appears moderate and notably less than for the raters in Menz et al. (2007). Is it possible that this may confound your estimates of the prevalence of OA at individual joints and differences in prevalence for different techniques?

Response: We have added the subheading ‘scorer reliability’

MM was the radiographer who conducted a training session with PMc the scorer in this study, but no inter-rater reliability study was undertaken. This is consistent with the discussion from the authors of the origin of the LFA that states “Consistent with all previous studies, however, inter-examiner reliability was somewhat lower, despite the level of training for the two examiners being identical. This suggests that there is some degree of inherent variability in the interpretation of some aspects of the atlas. Based on this observation, we concur with previous recommendations that, for research purposes, single examiners or consensus grading should be used to document OA from X-rays where possible” Menz et al 2007.

The OA scorer was MM and therefore to ensure better clarity in this we have removed the words ‘and OA scorer’.

Yes we agree that it is possible that our estimates of prevalence may have been confounded by the lower reproducibility of the rater in this study than that of the original authors of the LFA. There are a number of explanations for lower reliability scores in our preliminary work such as the positioning for the reliability study differed from that undertaken by the LTA, availability of only one view and lower quality of foot radiographs versus higher quality of resolution of electronic images used in year 23.

In addition, for the development and testing of the LTA, the same authors selected the radiographs for each LTA classification grade on which their reliability was calculated. This may provide more stable predictions of reliability scores but may not be as readily applicable to new raters external to the original development team.

We have amended the discussion of our limitations to reflect this.

At the individual joint level, our best levels of intra-rater reliability using the dorsoplantar view were shown to be in the 1st MTPJs which is consistent with LFA authors

We have removed this sentence from the manuscript
Results:

15. Can you please elaborate (in Discussion) how the prevalence of radiographic foot OA is greater for technique 3 than 1. I would have thought taking the more conservative option when scoring (some) items would lead to a reduction in prevalence of foot OA since you are ‘scoring down’. There appears to be inconsistency in use of hyphen versus en dash / em dash.

Response: We apologise for the error and are grateful to the reviewer for picking this up. Please see our response to point 13 above.

Discussion:

16. P9L226: '…discordance in case definition…' - can you please review this as to the reviewer's knowledge, there hasn't been any discordance in case definition of OA using the LFA.

Response: Thank you for highlighting this. We have removed the words “the discordance in case definition for” to avoid this confusion.

17. P9L232: Consider spelling out OARSI.

Response: this has been amended

18. P10L241: please insert a '.' after 'et al'. Please review entire Discussion.

Response: this has been amended.

19. P10L254: This paragraph could be omitted. You cant really compare your study to reference 10 since reference 10 had a definition of symptomatic OA (symptoms plus radiographs) whereas you have radiographs only so can only compare to studies that have published radiographic OA prevalence data.

Response: We do agree with the reviewer that we cannot directly compare our results to reference 10 as that investigation focussed on symptomatic foot OA. We were trying to make the point to the potential audience for the Journal of foot and ankle Research that they should be
aware of the two definitions and that the two definitions cannot be used interchangeably. We have amended the paragraph in anticipation that the point is more clearly articulated.

20. P11L265: A statement is made that the atlas is subjective because it is not interval/ratio scaled. Can you please review this sentence because 'subjective' is ambiguous. In addition, even if the atlas used a 100 point scale or more (and was interval/ratio which is not practical or necessary), it would still be 'subjective' in that it relies on human judgement for scoring.

Response: We were trying to suggest that atlases of radiographic OA use ordinal scales can be affected by subjective interpretation in comparison to an objective measurement (on an interval/ratio scale) that can be undertaken for assessing radiographic OA such as a measurement of joint space width. We appreciate that this was not clear in the manuscript and we have revised the text.

21. P11L271: Please consider omitting'…In addition, the normal joint space width reference used by the LFA is not based on a population study'.

Response: We have removed the sentence.

22. P11L275: The sentence regarding the reliability of the assessments doesn't not fit here where sensitivity across views is discussed.

23. Response: We have removed the sentence.

Limitations:

24. Please consider that you weren't able to obtain corresponding foot pain data on these participants to determine symptomatic foot OA prevalence.

Response: Yes we have incorporated consideration of this into the final paragraph of our discussion.

25. What are the implications of your moderate reproducibility in using the LFA on your results of radiographic prevalence across techniques?

Response:
As per point 14 above, we agree that it is possible that our estimates of prevalence may have been confounded by the lower reproducibility of the rater in this study than that of the original authors of the LFA. We have expanded the section of the first point in our limitations section that the reliability work was undertaken on an older sample of foot x-rays with only dorsoplantar views, prior to the data collection for the year 23 return cohort.

26. P12L294: the sentence regarding the reliability of the LFA versus the current study could be omitted. It really isn't the aim of the study.
Response: We agree and have removed

27. P12L301: Add in 'years' after '69-93'.
Response: this has been added

28. P12L306: Please consider rewriting this to remove mention of OA of the knee, hip or hands.
Response: We have altered the paragraph as suggested to

“The population that we investigated was all women. Whilst symptomatic foot OA may be more prevalent in women [10] it is possible that our prevalence estimates of foot OA may be higher than those for the populations that include both men and women”.

References:
29. Please review abbreviation of journal titles for references 6,8.
Please review the journal style for the bibliography (check if there is a need to write 'p' for page number).
Response: According to the JFAR author guidelines, an article within a journal should be written - All authors should be provided, The titles shouldn’t be italicised, No issue number is needed, No ‘p’ for page number is required, Year;vol:page no’s should have no spacing between. For example:


We have now updated the EndNote reference style to that of Biomed Central format that hosts the Journal of Foot and Ankle Research.
Tables:

30. Please check journal guidelines regarding presentation (e.g. no vertical lines).

Response: According to the JFAR guidelines for authors there is no specific request related to vertical lines. For presentation JFAR requires “Colour and shading may not be used. Parts of the table can be highlighted using superscript, numbering, lettering, symbols or bold text, the meaning of which should be explained in a table legend”.

We have changed the titles to reflect the JFAR guideline word count of 15.

Table 1:

31. Is there a need to report p-values for the Kappa values? Percentage agreement could be reduced to zero decimal points? Kappa values could be reduced to 1 decimal point?

Are the values showing dorso-plantar or lateral view assessments?

Response: We have amended kappa and agreement values so that they are clearer in the Table.

The values are from year 6 foot radiographs which are dorsoplantar view only. We have added dorsoplantar view onto the Table.

Table 2:

32. Please specify 'RG' or spell out.

A symbol associated with 'Number of ungradable joints' doesn't match the footer.

Please use superscript where appropriate within the footer.

Technique 1 - 2 values should be positive?

33. Response: We have written RG in full and amended the symbol and superscript within the footer.

Table 3:
Reviewer #2: This study examines different interpretive approaches for the LaTrobe Foot Atlas for scoring radiographic foot OA. Clarification is needed in several parts of the manuscript, particularly related to the sample and methods (see specific comments below).

1. **Abstract, Background:** The first two sentences of the Background do not seem to correspond to the aim of the study. What is the value of examining different interpretative approaches for an atlas that is not widely used?

Response: Yes, we agree this may confuse readers. In suggesting that the LFA has not yet been widely used, we meant that it hasn't been used by investigative teams that do not include one of the original authors. We believe that researchers investigating foot OA within the population do wish to use standardised approaches, of which the LFA is deemed the most appropriate at this current time. In this paper, we are advocating the use of LFA for such investigations but, based on our experiences, with the caveat that the interpretative scoring technique requires some further work to ensure concordance amongst international researchers. We have amended the abstract background accordingly.

2. **Abstract, Methods:** What was the mean age of the participants by this time point? Also, it should be stated in this section that radiographic OA was defined for the foot collectively and separately for individual joints.

Response: We have added the mean (SD) for age: 75.5 years (5.1)) and that radiographic OA was defined for the foot collectively and separately for individual joints.

3. **Background, page 5, lines 103-104:** Please provide values for reliability and validity.
Response: We have added the values as: (percentage agreement from 86-99% and weighted κ from 0.45-0.95).

4. Methods: A figure is needed to describe the flow of participants from the initial Chingford cohort to the current analytic sample. ([https://strobe-statement.org/index.php?id=strobe-home](https://strobe-statement.org/index.php?id=strobe-home))

Response: We have added Figure 1. Recruitment flow chart.

5. Methods, Radiographic Scoring of foot osteoarthritis: For Techniques 2 and 3, the authors describe the procedures for joints that were difficult to interpret and score. This also should be described for Technique 1.

Response: Thank you for highlighting this. We have added:

“Technique 1 was employed as the LFA standard technique [23] whereby all joints that were difficult to interpret and score for OPs and/or JSN were given a score based on a conservative estimate. (For example, where an OP in a participant’s joint may have been ambiguous to score between a grading of 2 or 3, the lower grading of ‘2’ was accepted).”

6. Methods, line 182-4: This sentence appears to be a result and should be included in the Results section.

Response: We understand the reviewers point, however the author team did discuss this and all authors believe that the preliminary reliability work was not a direct aim and thus result for this study but fits better for flow of thought as part of the underpinning methods for the scoring technique.

7. Methods, Statistics: Details on the statistical tests are needed.

Response: We have added “Intra-rater reliability was calculated by overall agreement (percentage of observed exact agreement) and weighted kappa statistic. Assessment of the different radiographic scoring techniques are described using frequency (%) of radiographic foot OA at person level and joint level. Differences between the techniques are reported as frequency range”.
8. Results: Please describe how the participants in this analysis (Visit 23 with foot radiographs) differed from the original cohort in terms of demographic and clinical characteristics.

Response: The participants at the year 23 visit were older, mean (SD) age: 75.5 (5.1) with a higher BMI, mean 27.6 (4.75) than at the baseline year 0 visit, mean (SD age: 54.05 (6.4) and BMI 25.3 (4.3).

We have added the values to the methods foot radiographs section.

9. Discussion, lines 239-243: More details about the sample for the study by Menz et al. (e.g., sample size, sex, age, geographic location) are needed to help the reader understand the comparisons being made between the study and the present analysis.

Response: We have added information from the Menz et al study as: “The sample size that Menz et al. (REF) investigated and age was similar to ours (n=198, mean age 75.9 years, [SD] 6.6), however they were drawn from a retirement village and a university health sciences clinic in North West Adelaide, Australia and 64% women whilst ours were all women drawn from a general population”.

10. Discussion, lines 289-291: Was the same positioning of participants used to obtain the foot radiographs from Year 6 as from Year 23?

Response: Thank you this is a good point raised by the reviewer. We have added the following sentences:

“There are a number of explanations for lower reliability scores in our preliminary work such as the foot positioning for the reliability study differed from that undertaken by the LTA, availability of only one view (non-weight-bearing dorso-plantar) and lower quality of foot radiographs versus higher quality of resolution of electronic images used in year 23”.

11. Discussion, lines 298-301: It is unclear how the differences in the cohorts are a limitation.

Response: To clarify we have rephrased the sentence to:

“There is currently no available foot radiographic data that compares different populations that may have different physiology, anatomy and genetics. As such, we do not know how representative the pictures used to explain the scoring method within the LFA are for global comparisons or how closely the Chingford 1000 Women’s study cohort foot radiographs may align to them”.

12. Discussion, line 308: The prevalence of radiographic foot OA is very high in this study. The authors should comment on whether this finding is clinically meaningful.

Response: Yes we agree and have added the following sentences:

“We are aware that the prevalence of foot OA was very high in our study. We believe this could be due to a combination of the population being all women aged 69 years in whom OA has generally been found to be more prevalent (van Saase 1989). Additionally OA was defined radiographically which has been shown to lead to higher estimates than other definitions such as self-reported and symptomatic (Pereira 2011). Estimates of the prevalence of OA of a similar order have been reported at other peripheral joints sites in other populations of older women (Yoshida 2002).”

13. Conclusion: The authors should elaborate on how this work in a single cohort using different interpretations of the same atlas “strengthens the case for harmonisation of case definitions.”

Response: In line with reviewer 1’s initial comments we have amended the wording of this in the conclusion of the main body of the manuscript as well as the abstract.

Reviewer #3: Dear Authors. This is an interesting article discussing the interpretive issues surrounding radiographic definitions of foot OA. I think the study would benefit from a greater detail in the discussion about the interpretation particularly the issues around certain anatomical sites and positional orientations that lead to the 3 techniques. I have specific comments and feedback, see below.

Introduction

1. Line 78 - no full stop.

Response: this has been amended

2. Lines 76 to 79 - please consider revising/removing the paragraph. Starting the article by discussing the disparities between symptomatic and radiographic OA is leading and suggests
that you are going to report PROMs in this article. It feels the article naturally starts at the next paragraph (Line 81).

Response: We appreciate your comment but we had feedback from Reviewer 1 (point 2) who asked for more detail to be added on the hand. In light of this we have decided to retain this first paragraph but have amended the sentence that raised the discrepancies between symptomatic and radiographic OA.

3. Line 106 - Please reference the sentence.

There are a number of interventional studies that have used the LFA to define foot OA:


Response: Thank you that is a good point. We actually meant that few other studies have been published that haven’t included a member of the original team that developed the atlas. We have amended this and included your reference suggestions.

Methods and Results

4. Line 177 to 184 Please can you be specific about whether there was only 1 scorer and was this PMc?

I cannot see the added value of table 1, intra rater reliability, unless the XRs were scored on multiple occasions. I would expect given the start of the paragraph that inter-rater reliability between the experienced and learning researcher would be more informative.

Response: Please note our response to reviewer 1, point 14 who also highlighted this.

If you would like to continue to include the intra-rater reliability, please can you add this to the discussion and compare the reliability LFA paper Menz et al. 2007:

Response: We understand the reviewers point, which is similar to reviewer 2s. The author team did discuss this and all authors believe that the preliminary reliability work was not a direct aim and thus result for this study but fits better for flow of thought as part of the underpinning methods for the scoring technique and therefore not to be a part of the discussion, but have now featured more discussion of this within our limitations section.

Discussion

It would have been helpful to include more observations on morphological differences between the alignments in this cohort and the LFA. Perhaps discussing whether difficulties leading to different techniques of XR interpretation are dependent on specific anatomical locations.

Response: Indeed, the other two reviewers also raised this point and we have now included discussion of these within the manuscript.