Author’s response to reviews

Title: Knee Pain and Related Health in the Community Study (KPIC): a cohort study protocol

Authors:

Gwen Fernandes (gwen.fernandes@nottingham.ac.uk)
Aliya Sarmanova (aliya.sarmanova@nottingham.ac.uk)
Sophie Warner (scw27@leicester.ac.uk)
Hollie Harvey (hollie.harvey@nottingham.ac.uk)
Kehinde Akin-Akinosoye (mbxka4@exmail.nottingham.ac.uk)
Helen Richardson (helen.richardson@nottingham.ac.uk)
Nadia Frowd (nadia.frowd@nottingham.ac.uk)
Laura Marshall (l.marshall@keele.ac.uk)
Joanne Stocks (mszjs2@exmail.nottingham.ac.uk)
Michelle Hall (michelle.hall@nottingham.ac.uk)
Ana Valdes (ana.valdes@nottingham.ac.uk)
David Walsh (david.walsh@nottingham.ac.uk)
Weiya Zhang (Weiya.Zhang@nottingham.ac.uk)
Michael Doherty (michael.doherty@nottingham.ac.uk)

Version: 2 Date: 22 Aug 2017

Author’s response to reviews:

Response Document

We would like to thank the reviewers and Editor for this opportunity to revise and improve our manuscript. We have made the necessary changes to the document (as highlighted (marked copy)). We trust we have met the suggestions the reviewers have made and look forward to hearing from you shortly.
Reviewer reports:

Niamh Moloney (Reviewer 1): Thank you for asking me to review this protocol for: Knee Pain and Related Health in the Community.

This is a comprehensive project and the authors are to be commended for the in-depth assessment that covers multiple dimensions of the pain experience as well as the planned long-term follow-ups. On the whole the project is well justified and described methods and analysis align well with project aims.

We thank you for this comment and for detailed consideration of our manuscript.

There are a few items that I think need some consideration.

Page 9 Lines 29-34

Can the authors describe in more detail what is meant by constitutional knee alignment (in early 20's) and how this is different to current knee alignment?

Can they also explain whether both are measured using the self-reported tool referenced and make it clearer that this a self-report measure?

1) This has been amended to include both a detailed description of constitutional alignment as well as specific mention of it being self-reported. (Methods, page 8, lines 219-224).

Page 10; Line 22-39
Why is a cut off of >/< 3 years used for early versus established knee pain? This appears incongruous with our standard definitions for acute, sub-acute and chronic pain. I can see that acute, sub-acute and chronic time frames may not be suitable for this study, however, calling <3 years early knee pain seems difficult to rationalize. Can the authors explain the rationale and supporting research evidence for using this time frame?

2) Participants were categorised as having ‘recent onset’ knee pain when they presented with symptoms for the first time in the past three years for most days of a one month period, unrelated to obvious major injuries. The time frame for this was initially set as KP onset in the past 1 year, however, due to difficulties in recruitment of ‘recent onset’ participants, this was extended to the past 3 years for any new KP symptoms.

Page 11: Please specify what will the urine and blood samples will be analysed

3) We have plans to measure urinary collagen type II crosslinks (CTX-II) which is a marker of collagen degradation predictive of osteoarthritis progression and serum levels of c-reactive protein (CRP), a marker of generalised inflammation. We also have plans to measure inflammatory markers related to pain specifically the pro-inflammatory cytokines IL-1, IL-6 and TNF alpha and the resolvin precursor 17-HDHA, which has been recently shown to correlate with severity of OA pain (Valdes et al Sci Rep 2017). We have now included this detail in the manuscript. (Methods, page 10, Lines 25-31).

Page 12, line 34: Please explain why balance will be assessed

4) This has been extended and explained with reference. (Methods, Page 13, Line 81-88).

Page 14: Why have only bony surfaces been selected for PPT testing? While I appreciate that joint line tenderness is characteristic in knee pain, and therefore makes sense in the context of the study, PPTs over bony margins are difficult to perform well eg controlling the algometer and hence the location and application rate. Most studies include soft tissue locations- either solely or in combination of joint lines. Further, sites such as tibialis anterior, or hand (or forearm) could be used as there are considerable reference data available for these sites. This would enhance confidence about whether subsequent results were within/outside normative values.

5) These sites were chosen to avoid influence of pain from other tissues, for example, muscle, ligaments and tendons. We agree that “muscle-deep pain” is widely addressed using QST approaches in studies of this nature, however, there is little experimental representation of
the “deep pain sensation” – a core characteristic of knee pain associated with OA (Cedraschi, C., et al., 2013). PPTs on such bony surfaces have been shown to be reproducible and recommended for experimental tests of evoked bone-associated pain (Andresen, T., et al., 2013).

Unmyelinated nociceptive afferents innervating the measured sites can allow assessment of evoked bone-associated pain (Andresen, T., et al., 2013). Thus, our approach will provide evidence specific to the “deep pain sensation” across localized, distal and remote sites in our study participants.

Page 16. A number of questionnaires will be added at the 1-year assessment on the basis that they will contribute to the psychosocial aspects of the individual's profile and may hold predictive value. I don't argue with this point but why are these not included at baseline and re-assessed at year 1? In my view this would make more sense, particularly if considering the predictive value of these measures, and understanding possible pain phenotypes associated with knee pain, incidence and deterioration. As this is the primary aim of the study, assessing these measures at baseline would seem important.

6) We thank the reviewer for this comment. At the baseline questionnaire, we were restricted by the number of questions to include in our survey covering a comprehensive spectrum of knee pain and known risk factors. For this reason we were unable to expand on the psychosocial aspects in the baseline questionnaire. However, for the Year 1 survey, as we were able to reduce some items from the baseline questionnaire, we were able to include more psychological assessments such as the life orientation test, the short item test on conscientiousness and an illness perception questionnaire. These questions will also be included in the Year 3 questionnaire. In addition to this, some measures such as the fibromyalgia score and the sleep questionnaire from the Medical Outcome Survey were regarded as important exposures that we did not consider during baseline discussions and in the baseline questionnaire. As a result of this, these surveys were subsequently included in the Year 1 and Year 3 questionnaire.

Sample size calculations; Thank you for presenting detailed sample size calculations for your measures. The only one that wasn't clear, or appears to be underpowered was PPT. Can the authors clarify that what they have included is correct (i.e. 400:100:100)?

7) The authors confirm that these figures are indeed correct for the three groups (400:100:100). However, the actual recruitment figure for the early knee pain group was 219 and this was due to difficulties in recruiting participants, e.g., who had early knee pain symptom development even in the past 3 years (see above point 2).
Discussion: Page 23 line 52. The authors explain that the method of knee pain classification is a standard method; however, here they refer to the method of recall for incident knee pain only, and not knee pain over the previous 3 years. How does this method relate to assessment of early versus established knee pain as outlined earlier in this manuscript? Further, it would be good to specify that this method of assessment relates to incident knee pain only.

8) Indeed, we used the current knee pain, defined as pain in or around knees for most days of the past month for the incident knee pain. This was different from early knee pain which was defined as knee pain (ever or current for most days of at least a month) in the past 3 years. We have now amended the statement to demonstrate that this definition was extended to three years for the current study recruitment and that it refers to incident knee pain only (Discussion, page 25, line 386-392).

Minor typographical errors:
Page 10, line 10 add 'be' after will in the sentence 'will then… identified'
Page 11, line 3 add 'be' after 'will then…obtained'

9) Many thanks! These typos have now been corrected.

Pim Luijsterburg (Reviewer 2): 2017-07-07

Thank you for the opportunity to review this study protocol with an topic of my interest. Please find below my comments:

- Page 7, line 27: Patients will be included of 40 years and over. However, at the ClinicalTrials.gov registration age 40-80 years is stated.

10) We thank the reviewer for identifying this discrepancy. We can confirm that we have now updated the clinicaltrials.gov website as there was no upper limit officially set for the study at baseline. Therefore, both the protocol and the clinicaltrials.gov registration page both state 40 years and over.

- Page 7, line 36: "Eligibility will be decided by the health professionals in each general practice" The design of this study is called a "general population cohort study". However, it seems that "eligibility will be decided by the health professionals in each general practice". Will they monitor who will not be invited to participate the study and inform regarding
differences with the invited persons (such as age, gender)? Who precisely is selecting eligible participants in the records of the general practitioner?

11) The reviewer is correct in stating that health professionals at each General Practice (GP) will determine suitable participants. Due to data protection legislation, the team at the University of Nottingham do not have the permission or consent to directly contact participants in the community. As a result, following successful ethical approval for this study, we worked via the Clinical Research Network in the East Midlands region to identify those GP practices (currently 12 in the Nottinghamshire and Derbyshire areas) who had the right resources available (time, manpower, logistics) to help screen suitable participants from the GP database who met our inclusion criteria (men and women aged 40 years and over, irrespective of knee pain status) and who also met our exclusion criteria (unsuitable participants were those who could not give informed consent, had a terminal or severe mental illness and those women who were pregnant (due to planned clinical assessments involving x-rays). Once these participants were identified, the GP practice would send each participant an introductory letter about the study and objectives as well as a questionnaire itself with a return envelope. 40,000 suitable participants were identified and 40,000 questionnaires were mailed out to the community. Out of these, 10,000 questionnaire responses were expected. The back page of the questionnaire also provided participants with the option of indicating whether they were willing to take part in further research on knee pain and knee osteoarthritis and an additional space to include their up to date contact details so that we could contact with them directly instead of going via the GP practices again.

- Page 20, line 17 and 41: What are the confounding factors?

12) We have included confounding factors that would be used for adjustment in our model. (Methods, page 21, Line 279).

- Page 20, line 36: Indeed cases and controls will not be fully matched. It is not clear if this was taken into account in the sample size calculation.

13) We have amended this to frequency matched by age and gender (the early knee pain group vs. established vs. no knee pain). (Methods, Page 21, line 285-286).

We did not consider confounding factors in the power calculation for all sub-studies nested in this cohort study. Apart from the frequency matching for age and gender, no other consideration in the power calculation for these sub-studies in this cohort. This caveat will be discussed in each sub-study when they get published as appropriate.
- Page 23, line 10: I am not sure if this cohort will represent the general population due to the eligibility selection of multiple health professionals.

14) We appreciate the reviewer’s comment on this point but hopefully have addressed this in our response 11. The health professionals working at the GP practices do not ‘select’ appropriate participants but rather ensure that the study inclusion and exclusion criteria are met before participants are sent a questionnaire via post.

- Page 24, line 2 (and figure 1): Here is stated n=400. In the ClinicalTrials.gov registration this is 450?

15) Thank you for identifying this. We have ensured that the clinicaltrials.gov website is now up to date with the correct figure of 400.

General: In the ClinicalTrials.gov registration is stated that study completion will be September 2017. Does this mean that the period of recruiting participants was a couple of years ago, or is it still ongoing? Publishing a study protocol should be before or during recruiting the targeted study population I suppose.

16) When the clinicaltrials.gov registration page was set up, the authors had confirmed funding for baseline (2014/2015) and Year 1 of the project (2015/2016). As the study then received further funding (March 2015) to carry on assessment of the cohort at Year 3, the official end date is anticipated to be September 2018. Further funding may be sought to study the cohort beyond Year 3.

Kind regards,

Pim Luijsterburg
Erik Lenguerrand (Reviewer 3): This is an interesting research but this protocol could be modified prior to submission:

- This research is aiming to investigate the natural history of knee pain using annual assessment over a period of three years. While the assessments are diverse and thorough, there is a risk that patients with transient and/or slower pain/OA "mechanisms" are not likely to be considered by this study. This is a limitation worse discussing.

17) We thank the reviewer for their comment and indeed acknowledge that there is a limitation in studying participants who do not meet the threshold for knee pain in the questionnaire time frame. For example, not having knee pain for most days of a one month period. However, we have used not only knee pain duration (past three years for recent onset and over three years for the established pain group) but also knee pain quantifiers such as whether pain is mild, moderate or severe using the numerical rating scale and knee pain descriptors such as intermittent or persistent using the Intermittent and Constant Pain (ICOAP) tool. We have made reference to this in the Discussion section as a caveat of misclassification bias. (Discussion, page 25, line 390-392).

- It is unclear whether patients with previous history of OA or arthroplasty in another joint will be recruited as well as patients receiving (pharmaceutical or not) treatment for pain in another part of their body. Such treatments is likely to influence the detection of patients with "no KP", "early KP"?

18) We agree with the reviewer that previous medical history of OA and any total or partial knee joint replacements are relevant to this questionnaire on knee pain. The questionnaire therefore included these questions as well as a section on all current medication that has either been prescribed or bought over the counter (including supplements, vitamins and alternative medicines). We also included a question on various treatments that participants have tried for knee pain (including drugs, exercises, changes to diet, footwear modifications, etc.). We have now included this in Methods, page 7, lines 192-197. It is true that such treatments would have an effect on symptom presentation and we have queried about knee pain for most days of a one month period over the past three years (in term of onset) but also this could be factored into subsequent analysis as a confounder at a later stage. We were not able to include further questions on any other joint replacements (hips, ankles, etc.) as these were additional questions and there was finite space in the baseline questionnaire.

- Similarly the role of other comorbidities and level of physical activities prior and after the study could confound the planned analyses.
19) We agree with the reviewer that comorbidities and level of physical activity may be confounders for the planned analyses. We have taken into account comorbidities by asking participants if they have ever been diagnosed with other conditions such as hypertension, diabetes, strokes, fibromyalgia, cancer and this section also included an open text box so that participants could list all comorbidities. We have included this in the protocol (Methods, page 8, line 213-215). The level of physical activity would be a confounder and whilst in the questionnaire, we included the SF 12 which has a physical component score (page 9, line 234-235), in the sub-study participants who attend for clinical assessments, we will use the GPPAQ (General Practice Physical Activity Questionnaire) to help quantify current levels of physical activity (Methods, page 11, lines 42-45).

Long-term participation in a longitudinal study is likely to be differential from one group of participant to another. In this specific study the loss to follow-up rate could differ by age, number and severity of comorbidities, severity of pain and function symptoms which could trigger an arthroplasty, or on the contrary limited interest in the research question in participants with no or minor KP symptoms at baseline. The resulting differential bias is likely to affect the investigation. Is there any strategy in place to maximise the participant/response rate in the course of the follow-up? If the current study design is not accommodating such a challenge, the authors could consider to use an imputation and/or weighting strategy during the analyses to reduce this bias.

20) We thank the reviewer for this insightful and thoughtful comment. It is indeed a bias that could exist in the current study. The strategy to retain participants throughout the course of the KPIC cohort timeline involves sending reminders for questionnaires as well as providing annual newsletters on the study progress in terms of recruitment numbers and preliminary results. However, we would consider using an imputation model or weighting strategy during the analyses to reduce the effects of differential bias. This has been included in our discussion section (page 26, line 399-402) and in the Methods (Statistical Analysis, page 21, lines 290-293).

- Please provide more details on the GP register as the international audience might not be familiar with this data source.

21) A GP register is a log of patients living in the local area of the GP surgery who are registered and may consult a General Practitioner without charge (free primary care service). Not all patients may be eligible for free secondary care (hospital care) services, but the GP register consists of people who reside locally including refugees, overseas visitors, students
and homeless people. Patients may be permanent or temporary depending on the length of
time they reside in an area (minimum of 24 hours). GP’s are also under duty to provide
emergency or immediate necessary treatment, where clinically necessary, irrespective of
nationality or immigration status. (Methods, page 6, line 161-165).

- Lines 19-24 Baseline, questionnaire section the sentence "…40,000 adults..registers" could
be better located in the participants, inclusion section on the previous page.

22) This sentence has now been shorted and moved to the previous page. (Methods, page 6, line
166-167).

- It would be informative to add the list of occupations classified by level of risk for OA in an
appendix. It is unclear how unemployed and retired participants are going to be classified.

23) We thank the reviewer for this suggestion. We have now included a table in the appendix of
high risk occupations. All participants who are currently retired were assessed based on their
occupations during their working career. If occupations were all low risk, they were classed
as low OA risk. The same principle was used for those who stated that they were currently
unemployed. This was clarified in the Methods section, page 9, line 230-234. We have also
provided additional information in Appendix 1.

- In the participants section, methods "The questionnaire will be accompanied..." The
questionnaire to check patient eligibility?

24) The questionnaire referred to here is the baseline questionnaire with the spectrum of
questions relating to pain, medical and medication history and covering a comprehensive
spectrum of known risk factors for knee pain. Sentence amended to make it clear now,
Methods, page 6, line 166-167.

- In clinical assessment "…will then identified..." "…will then be identified..."

25) Thank you, this has now been amended.

- "Written informed consent will then obtained.." "Written informed consent will then be
obtained.."
26) Thank you, this has now been amended.

- "research metrologist" the term metrologist is an unusual terminology.

27) This has now been amended to research professional as those collecting data as part of the study could be from varying backgrounds, allied health professionals, nursing, occupational therapy, biomedical scientist, etc. Hence the broad term, research professional, is perhaps the most appropriate to use.

- "in a standard fashion". I would advise to remove this term or to cite a guideline as the standard may differ from one place to another.

28) This has now been removed.

- Ultra sound: " the same equipment…were used…was performed..only one value per joint was recorded…” would it be better to use the present or future tense?

29) A future tense has now been used consistently.

- Three areas will be scanned with US. Will the same areas be systematically scanned for all patients?

30) Yes, this is correct. Line amended to reflect this (Methods, page 14, line 102).

- Gait analysis. Please add some references. The following sentence "An average of 6 gait cycles …and averaged for final analysis" Is it standard practice? Would it be better to discard the first cycle(s) and final cycle? A reference would help to have confidence in the proposed modelling strategy.

31) Thank you for this comment. The standard practice is anywhere from 3-6 gait cycles at a natural comfortable pace (Allali et al., 2015; Martin et al., 2013; Hass et al., 2012; Bilney et al., 2003). The participants will be directed to a well-lit environment and will be able to walk in their own footwear. Once a verbal cue was provided to begin walking, the participants will not be given any other cues and no feedback regarding their walking performance. We have now updated this section and included a reference for this section. (Methods, page 14, lines 109-113).
- the QST tests seem to be conducted at baseline and year 1 in the text but baseline and year 3 in the table.

32) QST will be conducted at baseline and Year 3. This has now been amended in the table.

- Please provide a reference to justify the calculation of the TS and mechanical sensitivity.

33) This methodology was replicated/adapted from the standardized protocol developed by the German Research Network on Neuropathic pain (Rolk et al 2006). This technique has been used across other experimental pain studies across knee pain populations (Hochman et al 2013). These references have now been included in the paper.

- "body pain mannequin" It could be informative to add this graphical representation in an appendix.

34) Thank you for this suggestion. We have now added the modified body pain mannequin as Appendix 2.

- It seems that there is two groups of "early KP", those identified at baseline and in the course of the follow-up. Are the same clinical assessments going to be performed on these two groups? Are they going to be analysed simultaneously? In this case, the benefit of the early matching is unclear.

35) This is correct. There are, in fact, three groups of ‘early KP’ participants identified at the baseline, Year 1 and Year 3 time points. The same clinical assessments will be conducted in every ‘early KP’ participant. Once recruitment has completed in September 2018, this group from the three time points may be combined and analysed simultaneously. However, the purpose of early matching at baseline was for the purpose of establishing differences between the ‘early KP’ group and the established and no KP group as part of a cross-sectional sub-study on identifying differences in pressure pain thresholds (PPTs) between the three groups.

- Figure 1 should stand alone. Please explain the "+" and "-", "++", "+++" symbols". In the "questionnaire part" of this figure it seems that patients with "Knee pain -" at Year 1, and then with "knee pain +" at year 2" are not followed at year 3 (to become Knee ++), as it is done with those in the left hand side harm of this graph? Similarly, in the "assessment part", participants in the Knee - and then Knee + are not followed at year 3 in the similar way as the
group with knee pain <= 3 years. Could it be justified in the text? I am also wondering whether the fact that patients with KP>=3y have not received further assessments at y2 and y3 deserves some explanations in the text.

36) We thank the reviewer for identifying this. There was indeed an error in the diagram as all participants who have consented to taking part in further research will be sent a follow up questionnaire (questionnaire arm of the Figure). With regards to the assessment arm, the focus will be on the early or incident knee pain participants. As the established knee pain and no knee pain groups have already been well studied in the literature, the focus due to the study objectives as well as logistics, resources and finance was on those with incident knee pain. This was made clear in the Figure (including any person at baseline who had no knee pain but then changed status to knee pain positive would be assessed) and in Methods, page 18, line 205-206.

The multiple + signs were removed and the Figure made simple with Knee pain positive (+) or knee pain negative (-) symbols.

- The section on sample size could be reformatted. It is quite long and difficult to read. Studies have generally a primary research question, a series of secondary ones, and are powered to address the main research question. If this is the case here, all the emphasis could be put on addressing the main research question:

"A sample of XXX will allow to detect a difference of YY in the primary outcome at a power of 90% (alpha=0.05, two-sided). This sample will also allow to test a difference of X points (ref) in secondary outcome XXX and of T points(ref) in outcome NNN…"

If all research questions are deemed equally important for this study and the authors want to stick to their current presentation, one might argue that the resulting multiplicity of statistical hypothesis tests ought to be accounted for in the derivation of the required sample sizes.

The authors could also simply provide one calculation on an outcome derived from the questionnaire to justify n=10,000; and another calculation on an outcome derived from the clinical assessment to justify n=400.

37) We appreciate the reviewer’s comments on the sample size calculation presentations. We have therefore focused the primary sample size calculations (incidence, progression and risk factors of knee pain) as suggested (page 18, lines 222-228), and listed three major sub-studies as we had thought about for information only.
However, the concepts of incidence and prevalence are used interchangeably throughout the manuscript. A logistic regression cannot be used to model incidence rate. If the authors are keen to keep their current presentation of the sample size section, point 4 on risk factors needs to be reconsidered. Point 3 also needs to be revisited as the underlying tool used for a sample size calculation is different if one wants to detect a difference in two rates or a difference in two rate changes. More importantly, the sample size based on "incidence of KP" (prevalence? of KP) is unclear (point 2). Are the authors aiming to determine the required sample size to assess a prevalence of 3% within a specific precision in the estimates of this prevalence? The current presentation is unclear and seems to be more relevant to the comparison of two prevalence rates but without the mention to what the 3% would be compared to.

Saying that a sample of 10,000+ questionnaires should be really more than enough for all the analyses based on the questionnaire outcomes!

38) Thanks for this comments. We agree that ideally a cohort study should be powered for a time-to-event outcome. However, as we do not know how long we will be able to follow up this cohort (funding is only secured for year one and 3 follow-ups!), and there is really no need to follow up knee pain incidence and progression every year. We therefore use dichotomous model and the annual incidence/progression from the general population for knee pain, and the odd ratio of 2 for a well-established risk factor overweight/obese to calculate the sample size. Logistic regression model (a well-established dichotomous model for common condition such as knee pain is therefore selected to estimate the power and for the future analysis. We believe this is adequate.

With regards to the analyses based on clinical assessments, it is unclear why a sample of 600 participants is required to conduct the analyses on PPT but only 400 participants will be recruited to this part of the study. Is it a typo? This paragraph could be better phrased especially the last sentence: "this unbalanced design…".

The section on US could be shortened.

39) The section has been revisited and the section on US shortened. There is no typo in the section on PPT sample sizes. The numbers needed in the groups were 400:100:100. However, as the study has already commenced, we now know that we were unable to recruit 400 early or incident KP participants due to various difficulties in identifying participants who only presented with recent onset KP. As the Editor has advised us to write this paper using the future tense, it is most appropriate to discuss the caveats of this sub-optimal recruitment in subsequent papers with the results of the study instead of this protocol paper.
This study is based on a cohort study design and the etiological/causal strength of its results is likely to be limited by the biases associated with most observational design. The findings are therefore more "descriptive" than causal and the question of power calculation is less important than for a trial, i.e. the question of exposing a larger group of participants than necessary to a new potentially inefficient/dangerous treatment is irrelevant to the current context and the merit of a power calculation of lesser interest in this context. I would therefore recommend to keep this section short.

40) Thank you for your comment. We trust that our amendments to this section are in keeping with your suggestions.

- The statistical analysis section could be better presented. This section is rather unbalanced: The section on "reproducibility of assessments" is too long and the one on SEM is disproportionate compare to the details proposed for the methods used to address objectives 1 and 5.

41) We have now reduced the section on reproducibility and moved the SEM discussion further up in the section as addressing the objectives of the cohort study.

While I believe the authors have a clear idea of what they want to do, the current format of the statistical analysis plan does not reflect a clear strategy and comes across as a "catch-all" section. It might be difficult at this stage to have an exhaustive view of all future analyses and better to only present a subset of representative analyses to explain the general strategy which will be used. When dealing with baseline measurements, a logistic, multinomial, or anova/linear regression will be fine. However, for analyses requiring to model the follow-up measurements, generalised linear mixed models or survival analyses will be required to account for the time participants will have been at risk, i.e "person-time" ( a simple logistic regression would be inadequate). The analyses based on the clinical assessments seem more complex with some based on participants only assessed once, on participants assessed several times or on participants matched to "control" participants with no KP.

The analyses might differ from one research question to another and one outcome to another. On top of the issue associated with repeated measurements and longitudinal follow-up, the clustering within surgery will need to be explored and if necessary accounted for in any analysis.

I would therefore encourage the authors to structure the presentation of their "broad" analysis plan around the 5 research questions listed in the objective sections.

They could list the key outcomes which would be investigated in each research question with a description of the main statistical framework which could be used and potential adjustments.
Ultimately, it will be easier to assess the relevance of the proposed statistical plan and more readable for the lay audience.

42) This section has now been re-structured with a focus on baseline analysis (first paragraph), follow up analysis (second paragraph), analysis of phenotypes of knee pain using the exploratory structural equation model (third paragraph) and reliability measures of clinical assessments (final paragraph).

With regards to the discussion section, some of the limitations mentioned above could be discussed. While I do agree with the risk of table 2 fallacy bias, there are other more or as important biases which could be mentioned in particular those inherent to an observational study design. Unless further explanations are provided to help the readership of the journal, I recommend to remove this comment and focus on other limitations more relevant to a clinical readership such as the selection bias associated to a potential differential loss to follow-up in the group with severe pain or frail participants.

43) The limitations section of this paper has now been expanded to include selection bias and what might be undertaken to address this (measuring responders and non-responders). Discussion, page 26, lines 399-402.