PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The Effect of Vitamin D Supplementation on Pain and Physical Function in Patients with Knee Osteoarthritis (OA) An OA Trial Bank protocol for a systematic review and individual patient data (IPD) meta-analysis
AUTHORS	Jin, Xingzhong; Antony, Benny; Wang, Xia; Persson, Monica SM; McAlindon, Timothy; Arden, Nigel K; Srivastava, Sudeepti; Srivastava, Rajeshwar; van Middelkoop, Marienke; Bierma-Zeinstra, Sita M.A.; Zhang, Weiya; Cicuttini, Flavia; Ding, Changhai

VERSION 1 – REVIEW

REVIEWER	Dr. (Prof.) RAJU VAISHYA Indraprastha Apollo Hospitals Sarita vihar, New Delhi INDIA
REVIEW RETURNED	03-Nov-2019

GENERAL COMMENTS	This systematic review has plenty of statistical inputs, for which I am not a specialist and hence unable to do justice in reviewing this manuscript.

REVIEWER	Jaiben George AIIMS, New Delhi, India
REVIEW RETURNED	23-Nov-2019

GENERAL COMMENTS	This is a protocol for systematic review and individual patient data metanalysis of randomized controlled trials assessing the effect of oral Vitamin D supplementation on pain and function in osteoarthritis of knee. The protocol describes in detail the methodology including the search criteria, data collection, risk of bias assessment, outcomes, and statistical analysis. It is well written and will be a valuable addition to the current literature. Comments: Page 7, Line 8-“As a minimum, included studies should have measured knee pain, physical function, serum levels of vitamin D, and basic patient characteristics including age, gender, and body mass index (BMI) at baseline.” – Please elaborate on serum Vit D levels as an inclusion criterion. Is it the levels at baseline, or after treatment? Vit D levels may be included for subgroup analysis, but having them as a requirement might decrease the number of studies.

REVIEWER	Prof Adrian Esterman University of South Australia, Australia

REVIEW RETURNED	04-Dec-2019

GENERAL COMMENTS	Well written paper. No issues with proposed statistical methods.

REVIEWER	Dr. George A. Kelley West Virginia University Morgantown, WV, USA
REVIEW RETURNED	05-Dec-2019

GENERAL COMMENTS	GENERAL COMMENTS Thank you for the opportunity to review this protocol aimed at conducting a systematic review and individual patient data (IPD) meta-analysis on the effects of vitamin D supplementation on pain and physical function in patients with knee osteoarthritis (KOA). Overall, I think this is a very important topic to address and that the authors have produced a very good initial protocol. To help strengthen this work, I have the following specific comments that are listed below. I am hopeful that these comments are taken in the spirit of producing the strongest IPD meta-analysis that is possible. SPECIFIC COMMENTS * Page 2, line 10 (Abstract) – “Randomized clinical trials” or “randomized controlled trials”? * Page 2, lines 27 through 31 (Abstract) – What is the rationale for these cutpoints? Why not just treat duration as a continuous variable? Also, as written in the Methods and analysis section of the paper, please add adverse events as a secondary outcome. * Page 2, lines 33 through 38 (Abstract) – This is a nice list of potential effect modifiers to examine. However, I would suggest that others also be added. These include the physical activity levels of the patients, dose of vitamin D ingested during the interventions, sleep, fatigue, anxiety, overall nutritional status, years that patients have had KOA, and years since the diagnosis of KOA. * Page 4, lines 5 and 6 (Introduction) – Suggest combining these first two sentences. * Page 4, line 53 (Introduction) – Suggest deleting the “a” before “there” and replacing “of” with “for”. * Page 4, lines 55 through 60 (Introduction) – Are there any systematic reviews of previous systematic reviews, with or without meta-analysis, on vitamin D supplementation and its effects on pain and physical function? If so, then please cite and briefly describe. If none exist, then please state such. * Page 5, lines 11 through 13 (Introduction) – Please reference the studies from which these post hoc analyses derived. * Page 5, last sentence of Introduction – As currently written, this would apply to both aggregate and IPD meta-analyses. Therefore, I would suggest that you rewrite this to reflect the important fact that an IPD meta-analysis, as opposed to an aggregate data meta-analysis, can examine the effect of potential effect modifiers, for example age, at the patient versus group level.

	* Page 5, line 19 (Methods and analysis) – After the first paragraph, I would suggest that you reorganize this section in the following order and with the following major subheadings: (1) Study eligibility, (2) Data sources, (3) Study selection, (4) Data abstraction, (5) Risk of bias assessment, and (6) Data synthesis. * Page 5, line 33 (Methods and analysis) – Why 1990 versus 1980, 1970, 1995, etc. as the start date? Also, since November 2019 has passed, why not something more recent, for example, December 31, 2019? Also, please tell the reader who will be responsible for conducting the searches. * Page 5, line 44 (Methods and analysis) – Please tell the reader whether duplicates will be removed electronically, manually, or both? * Page 5, last sentence – What about ClinicalTrials.gov? * Page 6, line 22 (Type of studies) – First, back in the abstract you defined RCT's as randomized clinical trials. Don't you mean randomized controlled trials? If the former, this means that you would include randomized trials without a control group, for example, a randomized trial that includes patients who received two different doses of vitamin D but in which no control group was included. Second, please tell the reader whether you're limiting your studies to those in which patient-level randomization occurred versus patient or group randomization. Third, are you including full-length studied published in journals or also including studies such as abstracts from conference proceedings, master's theses, and dissertations? Fourth, how will you handle studies that meet your inclusion criteria but have been retracted? One way to do this is to search a database such as Web of Science using the indexed term "retracted" for all studies that you plan to include, You could then delete any retracted studies from your final included database of studies or provide a strong rationale for not doing so (see: Fanelli D, Moher D. What difference do retractions make? An estimate of the epistemic impact of retractions on recent meta-analyses. bioRxiv. 2019:734137). Finally, I would suggest that you say that you will include as a supplementary file a reference list of all excluded studies, except duplicates, including the reasons(s) for exclusion after each citation. * Page 6, line 31 (Participants) – Please provide a minimum starting age here, for example, adults 18 years of age and older. * Page 6, lines 56 and 57 (Outcomes) – Please rewrite this sentence so that it is grammatically correct. * Page 7, line 55 (Risk of bias assessment) – Since the Cochrane instrument is used to assess risk of bias versus study quality, suggest that you replace "study quality" with "risk of bias" here and throughout the rest of the manuscript. * Page 8, line 8 (IPD collection and transfer) – When you say collaborate, please expand as to what this means. For example, will they be listed as co-authors on any work derived from this study, listed in the Acknowledgements, etc.?

	* Page 8, line 14 (IPD collection and transfer) – Replace “co-author” with “co-authors”. * Page 8, lines 15 through 19 (IPD collection and transfer) – I think an Appendix that includes a copy of this data delivery agreement would be helpful here. * Page 8, lines 48 and 49 (Variables of interest, Primary outcome variables) – As previously mentioned, what is the rationale for this 12-month cutpoint? Why not just treat duration as a continuous variable? * Page 9, lines 4 through 7 (Variables of interest, Primary outcome variables) – It appears that you’re limiting your physical function outcomes to self-reported versus performance instruments. If so, then please state such. * Page 9, lines 13 through 15 (Secondary outcome variables) – Again, why medium and short-term outcomes versus treating this as a continuous variable? * Page 9, line 29 (Potential treatment effect moderators) – As previously mentioned, suggest that you also include physical activity levels of the patients, dose of vitamin D ingested during the interventions, sleep, fatigue, anxiety, overall nutritional status, years that patients have had KOA, and years since the diagnosis of KOA as potential effect modifiers. * Page 10, lines 19 through 21 (Statistical analysis) – Please tell the reader what specific R packages you plan to use to conduct your analyses. * Page 10, line 25 (Conventional meta-analysis) – Since multiple random-effects models exist, please reference the one that you plan to use to conduct your aggregate data meta-analysis. Along those lines, the use of random-effects models have recently been challenged as more robust models have recently been proposed (see for example: 1. Doi SA, Barendregt JJ, Khan S, Thalib L, Williams GM. Advances in the meta-analysis of heterogeneous clinical trials I: The inverse variance heterogeneity model. Contemp Clin Trials. 2015;45(Pt A):130-138; 2. Doi SA, Barendregt JJ, Khan S, Thalib L, Williams GM. Advances in the meta-analysis of heterogeneous clinical trials II: The quality effects model. Contemp Clin Trials. 2015;45(Pt A):123-129; 3. Doi SAR, Furuya-Kanamori L, Thalib L, Barendregt JJ. Meta-analysis in evidence-based healthcare: a paradigm shift away from random effects is overdue. Int J Evid Based Healthc. 2017;15(4):152-160). * Page 10, lines 31 through 39 (Conventional meta-analysis) – Here and throughout, I think that it’s important to distinguish between heterogeneity (chi-squared, i.e., Q) and inconsistency (I-squared, an extension of Q, assessed as a percentage). Armed with this information, please rewrite to reflect this information and refer to the following, if necessary, for additional information: see: Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-560). Also, please provide a reference for your 30% and 50% cutpoints for “moderate” and “substantial” inconsistency.

	* Page 10, lines 42 through 48 (Conventional meta-analysis) – Since the time that the recommendation of Sterne et al. (2011), was published, more intuitive graphical and robust quantitative methods have been proposed for examining small-study effects. For example, see: Furuya-Kanamori L, Barendregt JJ, Doi SAR. A new improved graphical and quantitative method for detecting bias in meta-analysis. Int J Evid Based Healthc. 2018;16(4):195-203. * Page 11, lines 16 through 19 (IPD meta-analysis) – In addition to $p < 0.05$, I think it's important to also provide 95% confidence intervals. Also, for further thought, I would refer you to a recent series of articles regarding the use of statistical significance and $p < 0.05$ in the March 2019 issue of the American Statistician as well as a brief article in Nature about this topic (see: Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature. 2019;567:305-307). * Page 11, line 23 (IPD meta-analysis) – Replace "assumption" with "assumptions". * Page 11, lines 48 through 53 (One-stage modeling) – It appears that you are including one potential moderator in each of your models, which is fine. However, for those moderators in which a statistically significant association exists, what about building a larger model that incorporates multiple moderators to try and parse out truer associations. This model could be built by examining for multicollinearity between covariates ($r > 0.80$) and if multicollinearity exists, including the one variable that you consider to be the most clinically relevant. If both are considered equally relevant, you could include the one that traditionally has less measurement error. I would also consider doing the same for the two-stage modeling you've described. * Page 12, line 10 (two-stage modeling) – Why fixed versus random-effects here? * Pages 12 and 13 (Discussion) – Please make sure to include references when making relevant statements that require such. * Page 12, line 42 (Discussion) – To add strength to your statement about meta-analysis being an integrated part of precision medicine, you may want to mention the work of Gurevitch et al. (2018) as meta-analysis being the "grandmother of big data" (see: Gurevitch J, Koricheva J, Nakagawa S, Stewart G. Meta-analysis and the science of research synthesis. Nature. 2018;555(7695):175-182). * Page 13, lines 35 through 42 (Strengths and limitations) – You talk about missing studies from your analysis. Another approach to take, and this is in relation to the statistical analysis of your data, is to incorporate aggregate and IPD in the same analysis (see for example: Riley RD, Lambert PC, Staessen JA, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. Stat Med. 2008;27(11):1870-1893). END OF REVIEW

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Comments: This systematic review has plenty of statistical inputs, for which I am not a specialist and hence unable to do justice in reviewing this manuscript.

We would like to thank the reviewer for his time.

Reviewer: 2

Comments: Page 7, Line 8-“As a minimum, included studies should have measured knee pain, physical function, serum levels of vitamin D, and basic patient characteristics including age, gender, and body mass index (BMI) at baseline.” – Please elaborate on serum Vit D levels as an inclusion criterion. Is it the levels at baseline, or after treatment? Vit D levels may be included for subgroup analysis, but having them as a requirement might decrease the number of studies.

The vitamin D referred here is the serum level at baseline. We have modified the wording to make this clear. Since we will search RCTs that evaluate vitamin D supplementation, we expect these RCTs would have measured vitamin D level at baseline.

Reviewer: 3

Comments: Well written paper. No issues with proposed statistical methods.

We would like to thank the reviewer for his time.

Reviewer: 4

Comments: Page 2, line 10 (Abstract) – “Randomized clinical trials” or “randomized controlled trials”?

We have changed to the correct wording "randomized controlled trial".

Page 2, lines 27 through 31 (Abstract) – What is the rationale for these cutpoints? Why not just treat duration as a continuous variable? Also, as written in the Methods and analysis section of the paper, please add adverse events as a secondary outcome.

These cut-offs are based on common practice in knee OA research. These cut-offs were used in other publications in the field1, 2.

References:

1. Charlesworth, et al. Osteoarthritis- a systematic review of long-term safety implications for osteoarthritis of the knee. BMC Musculoskelet Disord 20, 151 (2019)

2. van Middelkoop M, et al. The OA Trial Bank: meta-analysis of individual patient data from knee and hip osteoarthritis trials show that patients with severe pain exhibit greater benefit from intra-articular glucocorticoids. Osteoarthritis and Cartilage 2016;24(7):1143–52.

We do not use duration as a continuous variable, because we expect there will be RCTs with repeated measures in pain (e.g. a 12-month RCT, but also measured pain at 3 and 6 months.) In such case, we would be able to investigate both the short-term and long-term effects. We have added mentioning adverse event as an outcome.

Page 2, lines 33 through 38 (Abstract) – This is a nice list of potential effect modifiers to examine. However, I would suggest that others also be added. These include the physical activity levels of the patients, dose of vitamin D ingested during the interventions, sleep, fatigue, anxiety, overall nutritional status, years that patients have had KOA, and years since the diagnosis of KOA.

Thanks for the suggestions. We agree there are many other potential effects that are interesting to be examined. As a protocol, we had to limit the length of our list to those are highly relevant to the primary outcomes (e.g. pain and function) and are most likely to be reported in vitamin D trials for OA.

Page 4, lines 5 and 6 (Introduction) – Suggest combining these first two sentences.

We have made changes to the sentence according to the suggestion.

Page 4, line 53 (Introduction) – Suggest deleting the “a” before “there” and replacing “of” with “for”.

We have made changes according to the suggestion.

Page 4, lines 55 through 60 (Introduction) – Are there any systematic reviews of previous systematic reviews, with or without meta-analysis, on vitamin D supplementation and its effects on pain and physical function? If so, then please cite and briefly describe. If none exist, then please state such.

There were 2 systematic reviews conducted on vitamin D trials for OA. We have mentioned this in the last sentence on Page 3 (revised version).

Page 5, lines 11 through 13 (Introduction) – Please reference the studies from which these post hoc analyses derived.

We have cited the references as suggested.

Page 5, last sentence of Introduction – As currently written, this would apply to both aggregate and IPD meta-analyses. Therefore, I would suggest that you rewrite this to reflect the important fact that an IPD meta-analysis, as opposed to an aggregate data meta-analysis, can examine the effect of potential effect modifiers, for example age, at the patient versus group level.

We have made changes to the sentence as suggested

Page 5, line 19 (Methods and analysis) – After the first paragraph, I would suggest that you reorganize this section in the following order and with the following major subheadings: (1) Study eligibility, (2) Data sources, (3) Study selection, (4) Data abstraction, (5) Risk of bias assessment, and (6) Data synthesis.

We have re-ordered the section as suggested. As this is a IPD meta-analysis, we kept the 'IPD collection and transfer' section between 'risk of bias assessment' and 'statistical analysis' to reflect to procedure order.

Page 5, line 33 (Methods and analysis) – Why 1990 versus 1980, 1970, 1995, etc. as the start date? Also, since November 2019 has passed, why not something more recent, for example, December 31, 2019? Also, please tell the reader who will be responsible for conducting the searches.

The year 1990 was chosen as a cut-off to limit the search to more current studies, and a previous systematic review1 has verified that no major RCTs were conducted for Vitamin D and osteoarthritis.

Reference:
1. Cao Y, et al. Association between serum levels of 25-hydroxyvitamin D and osteoarthritis: a systematic review. Rheumatology 2013;ket132.

We submitted the manuscript in October 2019, therefore November 2019 was the most recent date. At this stage, we completed the literature search in November, but would be happy to run an update search for any new published data. We have changed the dates as suggested.

Page 5, line 44 (Methods and analysis) – Please tell the reader whether duplicates will be removed electronically, manually, or both?

Duplicates will be removed electronically and manually. We have changed the wording to make this clear.

Page 5, last sentence – What about ClinicalTrials.gov?

We do not search ClinicalTrials.gov because only published RCTs will be eligible for inclusion. RCTs that are unpublished or yet-to-be-published are unlikely to provide IPD for analysis.

Page 6, line 22 (Type of studies) – First, back in the abstract you defined RCT's as randomized clinical trials. Don't you mean randomized controlled trials? If the former, this means that you would include randomized trials without a control group, for example, a randomized trial that includes patients who received two different doses of vitamin D but in which no control group was included. Second, please tell the reader whether you're limiting your studies to those in which patient-level randomization occurred versus patient or group randomization. Third, are you including full-length studied published in journals or also including studies such as abstracts from conference proceedings, master's theses, and dissertations? Fourth, how will you handle studies that meet your inclusion criteria but have been retracted? One way to do this is to search a database such as Web of Science using the indexed term "retracted" for all studies that you plan to include, You could then delete any retracted studies from your final included database of studies or provide a strong rationale for not doing so (see: Fanelli D, Moher D. What difference do retractions make? An estimate of the epistemic impact of retractions on recent meta-analyses. bioRxiv. 2019:734137). Finally, I would suggest that you say that you will include as a supplementary file a reference list of all excluded studies, except duplicates, including the reasons(s) for exclusion after each citation.

Thanks for the suggestions. Please see below response to each point:
• This review will include 'randomised controlled trial'. We have corrected the wording in the abstract.

• Both individually randomised trials and cluster randomised trials will be eligible. We have added this in the text.

• This review will included published RCTs. Generally, conference proceedings and student theses are not regarded as publications, and IPD are usually not available. We have reworded the paragraph to make this clear.

• As mentioned in the 'Literature search' section, we will use a systematic review workflow platform called Covidence, which will be able to handle studies that are excluded at every step.

• The platform Covidence will also be able to provide a PRISMA flow diagram that elicit the the number of exclusions and the reasons at each phase of the systematic review.

Page 6, line 31 (Participants) – Please provide a minimum starting age here, for example, adults 18 years of age and older.

We do not limit the age of the participants, as it is reported the disease can occur as early as 20 years old. However, most patients with OA defined by the American College of Rheumatology are usually over 50 years of age.

Page 6, lines 56 and 57 (Outcomes) – Please rewrite this sentence so that it is grammatically correct.

We have made changes and corrected the grammar.

Page 7, line 55 (Risk of bias assessment) – Since the Cochrane instrument is used to assess risk of bias versus study quality, suggest that you replace “study quality” with “risk of bias” here and throughout the rest of the manuscript.

We have mades changes to the wording according to the suggestion.

Page 8, line 8 (IPD collection and transfer) – When you say collaborate, please expand as to what this means. For example, will they be listed as co-authors on any work derived from this study, listed in the Acknowledgements, etc.?

The data custodians of included trials will be invited to list as a co-author. The details of the arrangement will be written on a data delivery agreement, which will undergo legal review by the OA Trial Bank and the custodians' institute. We decided not to have this information in the protocol manuscript as it is not directly related to the research methodology.
Page 8, line 14 (IPD collection and transfer) – Replace “co-author” with “co-authors”.

We have made changes according to the suggestion.

Page 8, lines 15 through 19 (IPD collection and transfer) – I think an Appendix that includes a copy of this data delivery agreement would be helpful here.

As mentioned previously, we decided not to include this information in the protocol manuscript as it is not directly related to the research methodology.

Page 8, lines 48 and 49 (Variables of interest, Primary outcome variables) – As previously mentioned, what is the rationale for this 12-month cutpoint? Why not just treat duration as a continuous variable?

We have addressed this comment. Please refer the previous response.

Page 9, lines 4 through 7 (Variables of interest, Primary outcome variables) – It appears that you’re limiting your physical function outcomes to self-reported versus performance instruments. If so, then please state such.

We acknowledge that performance-based function tests are more objective measures of physical function. We only include self-reported physical function as an outcome, because the aim of this review is to clarify the effect of vitamin D supplementation on clinical symptoms in patients with knee OA.

Page 9, lines 13 through 15 (Secondary outcome variables) – Again, why medium and shortterm outcomes versus treating this as a continuous variable?

We have addressed this comment. Please refer the previous response.

Page 9, line 29 (Potential treatment effect moderators) – As previously mentioned, suggest that you also include physical activity levels of the patients, dose of vitamin D ingested during the interventions, sleep, fatigue, anxiety, overall nutritional status, years that patients have had KOA, and years since the diagnosis of KOA as potential effect modifiers.

We have addressed this comment. Please refer the previous response.

Page 10, lines 19 through 21 (Statistical analysis) – Please tell the reader what specific R packages you plan to use to conduct your analyses.

We will use 'metafor' package for the conventional meta-analysis and two-stage IPD metaanalysis, and 'lme4' package for the one-stage IPD meta-analysis. We have added this information to 'Data synthesis and statistical methods' on page 9 of the revised manuscript.

Page 10, line 25 (Conventional meta-analysis) – Since multiple random-effects models exist, please reference the one that you plan to use to conduct your aggregate data meta-analysis. Along those lines, the use of random-effects models have recently been challenged as more robust models have recently been proposed (see for example: 1. Doi SA, Barendregt JJ, Khan S, Thalib L, Williams GM. Advances in the meta-analysis of heterogeneous clinical trials I: The inverse variance heterogeneity model. Contemp Clin Trials. 2015;45(Pt A):130-138; 2. Doi SA, Barendregt JJ, Khan S, Thalib L, Williams GM. Advances in the meta-analysis of heterogeneous clinical trials II: The quality effects model. Contemp Clin Trials. 2015;45(Pt A):123-129; 3. Doi SAR, Furuya-Kanamori L, Thalib L, Barendregt JJ. Meta-analysis in evidence-based healthcare: a paradigm shift away from random effects is overdue. Int J Evid Based Healthc. 2017;15(4):152-160).

We have added the reference for the 'DerSimonian and Laird' method for the random-effects model, and would like to thank for updates on the methodology papers on meta-analysis. We acknowledge there are caveats associated with the random-effects model, however, it is currently most widely used in meta-analyses and most commonly accepted in the OA research community, so we think it is better to use the random-effects model for this metaanalysis.

Page 10, lines 31 through 39 (Conventional meta-analysis) – Here and throughout, I think that it's important to distinguish between heterogeneity (chi-squared, i.e., Q) and inconsistency (I-squared, an extension of Q, assessed as a percentage). Armed with this information, please rewrite to reflect this information and refer to the following, if necessary, for additional information: see: Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-560). Also, please provide a reference for your 30% and 50% cutpoints for "moderate" and "substantial" inconsistency.

Much appreciated for this comment. We have rewritten the paragraph to reflect the different utility of the χ2 test and I2 score. The 30% and 50% cut-points are suggested in the Cochrane handbook, we have added this as a reference.

Page 10, lines 42 through 48 (Conventional meta-analysis) – Since the time that the recommendation of Sterne et al. (2011), was published, more intuitive graphical and robust quantitative methods have been proposed for examining small-study effects. For example, see: Furuya-Kanamori L, Barendregt

JJ, Doi SAR. A new improved graphical and quantitative method for detecting bias in meta-analysis. Int J Evid Based Healthc. 2018;16(4):195-203.

Many thanks for this comment. We have rewritten the paragraph to use the Doi plot and LFK index.

Page 11, lines 16 through 19 (IPD meta-analysis) – In addition to $p < 0.05$, I think it's important to also provide 95% confidence intervals. Also, for further thought, I would refer you to a recent series of articles regarding the use of statistical significance and $p<0.05$ in the March 2019 issue of the American Statistician as well as a brief article in Nature about this topic (see: Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature. 2019;567:305-307).

We agree that analysis results should not be presented with p-value alone. We have made changes to the wording to indicate that we will provide 95% CI.

Page 11, line 23 (IPD meta-analysis) – Replace "assumption" with "assumptions".

We have edited according the suggestion.

Page 11, lines 48 through 53 (One-stage modeling) – It appears that you are including one potential moderator in each of your models, which is fine. However, for those moderators in which a statistically significant association exists, what about building a larger model that incorporates multiple moderators to try and parse out truer associations. This model could be built by examining for multicollinearity between covariates ($r >0.80$) and if multicollinearity exists, including the one variable that you consider to be the most clinically relevant. If both are considered equally relevant, you could include the one that traditionally has less measurement error. I would also consider doing the same for the two-stage modeling you've described.

Many thanks for this suggestion. We agree that it will important the parse out the true effect of a moderator should multiple significant associations exist. We have added a third model in both one-stage and two-stage modelling.

Page 12, line 10 (two-stage modeling) – Why fixed versus random-effects here?

It should be 'random-effects'. We have made the correction and cited the DerSimonian and Laird 1984 paper.

Pages 12 and 13 (Discussion) – Please make sure to include references when making relevant statements that require such.

We have updated the references in Discussion.

Page 12, line 42 (Discussion) – To add strength to your statement about meta-analysis being an integrated part of precision medicine, you may want to mention the work of Gurevitch et al. (2018) as meta-analysis being the "grandmother of big data" (see: Gurevitch J, Koricheva J, Nakagawa S, Stewart G. Meta-analysis and the science of research synthesis. Nature. 2018;555(7695):175-182).

Many thanks for recommending the reference. Love the statement about meta-analysis being the 'grandmother of big data'.

Page 13, lines 35 through 42 (Strengths and limitations) – You talk about missing studies from your analysis. Another approach to take, and this is in relation to the statistical analysis of your data, is to

incorporate aggregate and IPD in the same analysis (see for example: Riley RD, Lambert PC, Staessen JA, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. Stat Med. 2008;27(11):1870-1893).

Many thanks for the recommendation. We have added this information to 'Strengths and limitation' section.

VERSION 2 – REVIEW

REVIEWER	Jaiben George AIIMS, New Delhi
REVIEW RETURNED	24-Dec-2019

GENERAL COMMENTS	All the comments have been addressed.

REVIEWER	Dr. George A. Kelley. FACSM West Virginia University, USA
REVIEW RETURNED	13-Jan-2020

GENERAL COMMENTS	GENERAL COMMENTS Thank you for the opportunity to review this revised protocol aimed at conducting a systematic review and individual patient data (IPD) meta-analysis on the effects of vitamin D supplementation on pain and physical function in patients with knee osteoarthritis (KOA). Overall, I think the authors have a decent job in responding to my comments. My remaining and minor comments, are listed below. When commenting, I refer to line numbers in the unmarked manuscript. SPECIFIC COMMENTS * Page 4, last two lines of page 4 and first two lines of page 5 – There appears to be some confusion here. While I did see the references to the two previous systematic reviews in your original submission and again in this revised version, I was asking about systematic reviews of previous systematic reviews, with or without meta-analysis, in other words, a review of reviews. As an example of what I'm talking about, please see the following: Kelley, G. A. and K. S. Kelley (2015). "Meditative movement therapies and health-related quality-of-life in adults: A systematic review of meta-analyses." PloS One 10(6): e0129181. * Page 5, lines 44 and 45 – Since you're including cluster randomized controlled trials, you will probably need to consider how these will be handled statistically since they are different than a traditional randomized controlled trial, for example, standard errors. * Page 5, line 49 – Since you state in your responses to my comments that you will not limit the age of participants, I would suggest that you include this information in the manuscript itself. * Page 6, lines 35 and 36 – In your response, you provide what I believe to be an acceptable rationale for starting with the year 1990. However, since the reader may question why you chose this cut-point, I would suggest that you add this information to the manuscript itself.

	* Page 8, lines 17 through 43 (IPD collection and transfer) – You've stated in your response that you didn't include information about the data agreement because it wasn't associated with the research methodology. However, I actually think that is associated since the methods used to retrieve IPD can have an important impact on how much data is actually obtained. In addition, given that we are now in an age of reproducibility, not to be confused with replication, I think it's is important to include as a Supplementary file an example of what this agreement might look like. This will also be helpful to other researchers as they plan their own IPD meta-analyses. * Page 9, lines 22 and 24 – Based on your response, I'm fine with your use of the proposed cut-points for both short and long-term outcomes. However, to help the reader, I would suggest that you insert your response into the manuscript itself so the reader is aware of why you're doing what you're doing. Also, when pooling results, you'll probably need to think about how you will handle multiple assessments of the same outcome in the same participants over different time periods since they're not independent of each other. * Page 9, starting on line 38 (potential treatment effect moderators) – In addition to your already nice list of potential moderators, I had previously suggested that you consider adding the following variables as potential moderators: physical activity levels of the patients, dose of vitamin D ingested during the interventions, sleep, fatigue, anxiety, overall nutritional status, years that patients have had KOA, and years since the diagnosis of KOA. Your response was as follows: "We agree there are many other potential effects that are interesting to be examined. As a protocol, we had to limit the length of our list to those are highly relevant to the primary outcomes (e.g. pain and function) and are most likely to be reported in vitamin D trials for OA." With respect to the above, I believe that one of the most important reasons for conducting any type of systematic review, with or without meta-analysis, is to provide direction for future research. Thus, including the variables I suggested would serve two purposes. The first, and most obvious, would be the potential to look at the association between these variables with changes in pain and physical function. The second, but less obvious, is to provide direction for future researchers on what "should" be included versus what "is" included. * Page 13, first written line – First, you're welcome about the Gurevitch et al. quote. However, to not misquote Gurevitch et al., I would suggest that you start this sentence as follows: "Meta-analysis is considered by some…." Gurevitch et al., were actually referring to meta-analysis overall and not just an IPD meta-analysis. END OF REVIEW

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2

Comments: All the comments have been addressed.

We would like to thank the reviewer for spending time on the manuscript.

Reviewer: 4

Comments: * Page 4, last two lines of page 4 and first two lines of page 5 – There appears to be some confusion here. While I did see the references to the two previous systematic reviews in your original submission and again in this revised version, I was asking about systematic reviews of previous systematic reviews, with or without meta-analysis, in other words, a review of reviews. As an example of what I'm talking about, please see the following: Kelley, G. A. and K. S. Kelley (2015). "Meditative movement therapies and health-related quality-of-life in adults: A systematic review of meta-analyses." PloS One 10(6): e0129181.

We have done a literature search for reviews of systematic reviews on this topic and confirmed there are no published review of systematic review of vitamin D on osteoarthritis. We have added a statement to reflect this on page 5.

* Page 5, lines 44 and 45 – Since you're including cluster randomized controlled trials, you will probably need to consider how these will be handled statistically since they are different than a traditional randomized controlled trial, for example, standard errors.

Thanks for the suggestion. We planned to include cluster RCT in the review, however, our literature shows that there are no cluster randomised controlled trials on this topic. Regardless, we have added a short sentence in the statistical methods on page 10 to briefly mention how the clustering effect will be adjusted if a cluster RCT is included.

* Page 5, line 49 – Since you state in your responses to my comments that you will not limit the age of participants, I would suggest that you include this information in the manuscript itself.

We have added this information in the manuscript.

* Page 6, lines 35 and 36 – In your response, you provide what I believe to be an acceptable rationale for starting with the year 1990. However, since the reader may question why you chose this cut-point, I would suggest that you add this information to the manuscript itself.

We have added this information in the manuscript.

* Page 8, lines 17 through 43 (IPD collection and transfer) – You've stated in your response that you didn't include information about the data agreement because it wasn't associated with the research methodology. However, I actually think that is associated since the methods used to retrieve IPD can have an important impact on how much data is actually obtained. In addition, given that we are now in an age of reproducibility, not to be confused with replication, I think it's is important to include as a Supplementary file an example of what this agreement might look like. This will also be helpful to other researchers as they plan their own IPD metaanalyses.

An example of the data transfer agreement is publicly available on the OA Trial Bank website : https://www.oatrialbank.com/procedures/. We have added a sentence to direct the readers to the link should they are interested in looking at the agreement.

* Page 9, lines 22 and 24 – Based on your response, I'm fine with your use of the proposed cut-points for both short and long-term outcomes. However, to help the reader, I would suggest that you insert

your response into the manuscript itself so the reader is aware of why you're doing what you're doing. Also, when pooling results, you'll probably need to think about how you will handle multiple assessments of the same outcome in the same participants over different time periods since they're not independent of each other.

We have added a sentence in the 'Primary outcome variables' section to explain that these cutoffs were used in other systematic review in the field.

* Page 9, starting on line 38 (potential treatment effect moderators) – In addition to your already nice list of potential moderators, I had previously suggested that you consider adding the following variables as potential moderators: physical activity levels of the patients, dose of vitamin D ingested during the interventions, sleep, fatigue, anxiety, overall nutritional status, years that patients have had KOA, and years since the diagnosis of KOA. Your response was as follows:

"We agree there are many other potential effects that are interesting to be examined. As a protocol, we had to limit the length of our list to those are highly relevant to the primary outcomes (e.g. pain and function) and are most likely to be reported in vitamin D trials for OA."

With respect to the above, I believe that one of the most important reasons for conducting any type of systematic review, with or without meta-analysis, is to provide direction for future research. Thus, including the variables I suggested would serve two purposes. The first, and most obvious, would be the potential to look at the association between these variables with changes in pain and physical function. The second, but less obvious, is to provide direction for future researchers on what "should" be included versus what "is" included.

Thanks for the comment. We agree that it would be interesting to analyse the association between the suggested variables with changes in pain and physical function. However, we feel that the inclusion of these variables would somewhat deviate from the overarching aim of this project, which is to identify OA phenotypes that potentially benefit from vitamin D supplementation. While some of the suggested variables, such as physical activity, may have effect on knee pain and function, they are considered as lifestyle risk factorsrather than disease phenotypes. We selected the current list of moderators with the intention to provide direction in patient selection for future RCTs of vitamin D treatment in OA sub-groups. Therefore, after careful consideration, we decided not to include the suggested variables.

* Page 13, first written line – First, you're welcome about the Gurevitch et al. quote. However, to not misquote Gurevitch et al., I would suggest that you start this sentence as follows: "Metaanalysis is considered by some…." Gurevitch et al., were actually referring to meta-analysis overall and not just an IPD meta-analysis.

Thanks for the suggestion. We have revised the sentence accordingly.

VERSION 3 – REVIEW

REVIEWER	George A. Kelley West Virginia University USA
REVIEW RETURNED	03-Feb-2020

GENERAL COMMENTS	GENERAL COMMENTS Thank you for the opportunity to review this second revision of a protocol aimed at conducting a systematic review and individual patient data (IPD) meta-analysis on the effects of vitamin D

	supplementation on pain and physical function in patients with knee osteoarthritis (KOA). The authors have done a decent job in responding to my second set of comments. I only have minor clarifying comments and suggestions. When commenting, I refer to the marked manuscript. SPECIFIC COMMENTS * Page 5, line 5 – Thank you for addressing the issue of systematic reviews of previous systematic reviews, with or without meta-analysis. However, rather than make such a definitive and strong statement, and so it flows better with the previous sentence, I would suggest that you rewrite this as follows. “In addition, to the best of our knowledge, no previous systematic reviews of previous systematic reviews on this topic exist.” This statement more accurately reflects what is inherent in any search of the literature, the chance of missing potentially eligible studies. * Page 6, line 52, Replace “choose” with “chose” * Page 10, line 55 and 56 – If you’re going to use the Dersimonian and Laird model, I would strongly suggest that you say that you will use the Knapp-Hartung adjustment since it will always lead to better coverage than the non-adjusted value. Inthout (2014) describes this and shows how it can be easily implemented in Excel (see: IntHout, J., et al. (2014). "The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method." BMC Medical Research Methodology 14). The rationale that the Dersimonian and Laird model is “what is usually used in the field” is unacceptable. From my perspective, this speaks to the naivete of the field with respect to high-quality, objective, meta-analyses. Across all fields, I strongly suspect that the use of the Dersimonian and Laird model continues to be used because (1) it has existed since 1986, (2) is commonly included in most “canned” statistical packages, and (3) is carried out by those who have little knowledge regarding the most robust methods currently available for conducting meta-analysis. * Page 10, last written line, and first written line on page 11 – A better way to say this may be as follows: “If an included study is a cluster RCT, results will be corrected using previously established procedures.” * Page 13, line 30 – Insert “to be” after the word “some” * With respect to your response to my repeated suggestion about the addition of potential moderators – If you don’t think my suggestion about adding these potential modifiers will affect your results, I’m okay with excluding them. However, if you think they will affect your results, then they should probably be included. If the latter, addressing this issue now would be much better than having to redo your entire analysis when an external reviewer makes the same suggestion after the original manuscript is submitted. END OF REVIEW

VERSION 3 – AUTHOR RESPONSE

Reviewer: 4

Comments: * Page 5, line 5 – Thank you for addressing the issue of systematic reviews of previous systematic reviews, with or without meta-analysis. However, rather than make such a definitive and strong statement, and so it flows better with the previous sentence, I would suggest that you rewrite this as follows. “In addition, to the best of our knowledge, no previous systematic reviews of previous systematic reviews on this topic exist.” This statement more accurately reflects what is inherent in any search of the literature, the chance of missing potentially eligible studies.

We have rewritten the statement as suggested.

* Page 6, line 52, Replace “choose” with “chose” Thanks for spotting the grammatic error.

We have corrected the error as suggested.

* Page 10, line 55 and 56 – If you’re going to use the Dersimonian and Laird model, I would strongly suggest that you say that you will use the Knapp-Hartung adjustment since it will always lead to better coverage than the non-adjusted value. Inthout (2014) describes this and shows how it can be easily implemented in Excel (see: IntHout, J., et al. (2014). "The HartungKnapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method." BMC Medical Research Methodology 14). The rationale that the Dersimonian and Laird model is “what is usually used in the field” is unacceptable. From my perspective, this speaks to the naivete of the field with respect to high-quality, objective, meta-analyses. Across all fields, I strongly suspect that the use of the Dersimonian and Laird model continues to be used because (1) it has existed since 1986, (2) is commonly included in most “canned” statistical packages, and (3) is carried out by those who have little knowledge regarding the most robust methods currently available for conducting meta-analysis.

Thank you for this comment. We have carefully read the suggested literature comparing the HKSJ method and DL method. The HKSJ method appears to outperform DL method particularly in meta-analysis with a small number of trials, which is likely the case for this systematic review. Therefore, we have now changed to use the HKSJ method for our convent meta-analysis.

* Page 10, last written line, and first written line on page 11 – A better way to say this may be as follows: “If an included study is a cluster RCT, results will be corrected using previously established procedures.”

We have rewritten the sentence as suggested.

* Page 13, line 30 – Insert “to be” after the word “some”

We have made the changes as suggested.

* With respect to your response to my repeated suggestion about the addition of potential moderators – If you don’t think my suggestion about adding these potential modifiers will affect your results, I’m okay with excluding them. However, if you think they will affect your results, then they should probably be included. If the latter, addressing this issue now would be much better than having to redo your entire analysis when an external reviewer makes the same suggestion after the original manuscript is submitted.

Thanks for the comment. We had a discussion and believed adding these potential moderators will not affect the results, because as far as we know, most of the trials on this topic were very

simple trials which did not collect these variables at baseline. As mentioned in our previous response, we selected the current list of moderators with the aim to identify OA phenotypes that potentially benefit from vitamin D supplementation. These moderators are more likely to be reported in vitamin D for OA trials.