PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The effect of combined resistance exercise training and vitamin D3 supplementation on musculoskeletal health and function in older adults: A systematic review and meta-analysis
AUTHORS	Antoniak, Anneka; Greig, Carolyn

VERSION 1 - REVIEW

REVIEWER	Justin C. Brown Dana-Farber Cancer Institute, Boston MA, United States
REVIEW RETURNED	17-Oct-2016

GENERAL COMMENTS	This is an interesting and much needed review of studies published examining the effects of vitamin D supplementation and strength training among older adults. The justification, methods, and results are reasonably well-described. However, I feel the authors have missed an opportunity in the discussion to provide recommendations to substantially improve the quality and clinical utility of the research in this area. The major strengths of this paper include a pre-registered protocol, multiple data abstractors, and a clearly defined rationale for why this paper is needed. My enthusiasm for this paper is tempered by a few, mostly fixable, issues. The authors should consider the following: Clearly defining what efficacy comparison each of the two groups of trials tried to answer will help to quickly familiarize readers with what groups are being compared. I think group 1 is describing the additive benefit of vitamin D supplementation when combined with strength training. And I think group 2 is describing the additive benefit of strength training when combined with vitamin D supplementation. I think this specificity of phrasing in the abstract and main paper will help readers beyond what current text is used to describe the two groups of trials. It would be helpful if the authors could confirm that all of the studies in group 1 used a placebo control. In table 2, the exercise and vitamin D protocols are described, but the alternative control conditions are not, and this may have implications for interpreting the efficacy comparisons. My biggest concern with trials in the area is that many outcomes are measured and they are selectively reported. For example, DXA imaging can provide dozens of variables about muscle, fat, and bone. Could the authors comment on what was the pre-specified

	primary outcome declared on clinicaltrials.gov for each trial and was the primary endpoint was met? Did other outcomes happen to sneak into papers that weren't declared a priori as secondary or supportive outcomes? I worry that a lot of what is being reported may be false-positive signals, therefore discussing what were planned primary and secondary outcomes versus what was ultimately was reported will be of incredible value to the field. I would strongly discourage the authors from discussing or highlighting changes that were significant within only one group. For example, I think, in table 3 there are several outcomes with significant p values but make mention of "no between-group difference". The authors have the opportunity to highlight that significant within-group changes are often presented to salvage a group-by-time interaction that was not statistically significant in an RCT (see: Allison 2016 Obesity: Common Scientific and Statistical Errors in Obesity Research, as an example). I would limit the column in table 3 labeled significant results to only those results where a significant group by time effect was demonstrated as this is the unbiased efficacy comparison derived from an RCT. The authors mention that publication bias was not detected, but no formal statistical examination of publication bias was conducted, and it appears the search was limited to only [presumably] published "journal studies." It would be helpful if the authors could elaborate how this statement is supported, as it is very likely that null trials may not be published. If so, the lack of searching for non-published trials may mean that your efficacy estimates are biased away from the null and overestimate the effects of these interventions. The present study did demonstrate that select outcomes including lower-extremity muscle strength and other functional outcomes did improve in both groups of efficacy studies. If at all possible, I would encourage the authors comment on the clinical utility of what these changes mean for patients and providers. For example, is a 1 kg increase in muscle strength important for a sarcopenic older adult? This may be a challenging question to answer, which has implications for future studies in this area. I think the authors should provide recommendations as to what work needs to be done to definitively answer the lingering questions in this area. What study designs are useful (i.e., 2x2 factorial)? What endpoints are the most persuasive to patients and providers? Now that the authors have summarized the current data, they should provide guidance as to what they feel is the best approach moving forward. On a related note, the field of pharmaco-therapy for cancer cachexia is going through similar growing pains (see: Fearon 2015 J Cachexia Sarcopenia Muscle) and probably many lessons from cachexia with respect to study design can be applied to older adults in this setting. I believe that if we sincerely want exercise and supplementation to be utilized in routine clinical practice, then clinical trials in this area need to be held to the highest standards such as that with pharmacotherapy. I think, if the authors are amenable to my comments, shifting the focus of the discussion from elaborating on individual trials to the more global strengths and limitations of the body of research in the totality of its current form can provide an impetus to improve the conduct and clarity of reporting of RCTs in this area. Ultimately, this will help move the field forward and provide

	a convincing rationale as to why these interventions merit integration into routine clinical practice.

REVIEWER	Kirsti Uusi-Rasi The UKK Institute for Health Promotion Research, Finland
REVIEW RETURNED	26-Oct-2016

GENERAL COMMENTS	General comments The aim of this systematic review was to assess the additive or combined effect of resistance training and vitamin D supplementation on musculoskeletal health, and the topic is interesting, and the authors have done lot of work with analyses. They remark that there is lack of reviews evaluating the effects of vitamin D and exercise, but there is also an obvious reason for this; there are no trials to be included. Of note, there is one review/meta-analysis of Minshull et al (Calcif Tissue Int, 2016) which should be mentioned in this paper. However, the paper in its current form raises many questions and concern. The most important is that the paper should be easier to read. There are so many groupings and so many test that it is difficult to follow when you are speaking about exercise effects, and when vitamin D effects. Please, pay attention to the clarity of the text. The groups are clearly stated in the beginning of the Discussion, but elsewhere it is not easy to follow the text. The trials included in this meta-analysis are divided into two groups, and in most outcomes there are only two studies included. Group 1 compares vitamin D vs. placebo among exercisers. Group 2 compares exercise vs. non-exercise among participants who all were vitamin D supplemented. Neither of the groups included a pure control group with placebo and no exercise. This means that placebo or learning effect cannot be ruled out. Actually, there are only two studies which could answer to the above mentioned question (refs. 21 and 32). It would be very helpful for the readers, if you could describe the methodology a little bit more. You have most likely followed the Cochrane instructions, but a short review of the methods could be useful. For instance, how was taken into account the different study sizes, number of participants ranging from 20 to 409, and the duration of the intervention from 12 weeks to 24 months? Which variables influence on the weight percent used in the analysis? Other comments The title is not accurate. The analyses compare effects of vitamin D vs. placebo among exercisers, or exercise effects among vitamin D users, but no combined effects. Introduction is confusing. You have referred to several papers relating to sarcopenia, which is not the study question. You should concentrate in physical functioning. A normal age-related change in physical functioning or muscle strength is not sarcopenia. The conclusion is too strong. Based on this review it is not possible to speak about additive effect. Why the paper of Kukuljan et al. is not included? They had a 2x2

	factorial design with fortified milk with vitamin D and calcium, and exercise. It would be more useful for comparison with Agergaard et al, because these are the only studies including only men. Sex is probably more confounding factor than protein intake, at least in this case because protein intake was sufficient in all groups. Ref. 36 (Verschueren et al.) is not included in any of the analysis. Why not? Table 2 shows number of participants in each study, but the given numbers are those who completed the study, not the randomized participants. However, the trials have used intention-to-treat analyses. Include also the sex of the participants in the table. Please, open all abbreviations in the footnotes. Table 3: Please add the reference numbers after the name of the author and year. Page 12 and Summary (page 13): Please, be careful with the directions of the changes. At least Uusi-Rasi et al. (ref. 21) did not find an improved normal walking speed with vitamin D. On the contrary, walking speed deteriorated in all groups except the group with exercise without vitamin D. The same was true with 5 x chair stand time, which improved more in the exercise groups, and no change with vitamin D without exercise was found. Of note, in addition to ref 21, references 32 and 33 have used SPPB-test, which includes normal walking speed and 5 x chair stand test. The results should be available if requested. Bunout et al. (ref. 32) did not show an improvement in TUG-time in any of the groups. TUG-test time increased in all groups indicating that it took longer to do the test. Ref. 21 also showed declined TUG time in non-exercisers with vitamin D, and no change in exercisers with vitamin D. Figure 2: It is not obvious how the results are calculated. Bunout et al (ref. 32) showed actually a greater increase in quadriceps strength in the exercisers without vitamin D than in the exercisers with vitamin D. In the study ref. 21 the increase was similar in exercisers with and without vitamin D. A small study of Agergaard et al. (ref. 31) cannot have such a strong effect on results. That is why it is difficult to agree with the authors that “A greater increase of muscle strength in replete older adults represents a novel finding of this review”. Compared to what? (Page 13/14). These results only confirm the previous finding that exercise improves muscle strength. Maybe the best conclusion is that these studies were unsuitable for comparison (page 14). It is difficult to include these three studies in the same analysis, because all studies have used different methods. How did you calculate the mean (sd) for the muscle strength? They are not presented as such in the papers. Discussion for group 2 should include studies showing exercise effects (training vs. no training) on musculoskeletal health, because in this group all participants were supplemented with vitamin D. Now there is discussion about vitamin D effects, all studies (44-48) evaluated vitamin D, not exercise. The difference between exercisers vs. non-exercisers can be a learning effect, not vitamin D effect. Even if a vitamin D effect existed, this design could not show it, because all participants were supplemented. Page 15, narrative analysis: Too much attention is given to a muscle quality in narrative analysis, the only study using this outcome was

	the pilot study of Agergaard et al. (ref. 31), in which the number of older men was 7 vs. 10. Page 9, 1st and 2nd paragraphs: There is also an inaccuracy in hip BMD measurements. Most studies have reported femoral neck BMD, not hip BMD. See also figure 4 and 13. In summary: Please conclude the results separately for Group 1 and Group 2, preferably in different sections, to make it easier for readers to get the message. Table 4: Please, add the reference numbers in the column Author, year. Table 5: Please, use the same reference numbers of the studies throughout the paper. Figures 2-5, and figures 6-14: It is not obvious where the numbers come for the mean (SD) of the intervention or control groups. Could you please clarify? All studies do not give the exact end point values but the change from baseline to the end (adjusted change of percent). Have you calculated the means or requested the data from the authors? Keep also the lay-out similar in all tests; in all other figures the control is on the left and Intervention on the right, but TUG is shown vice versa. Table 7 and 8: Please give the description of the group 1 and group 2 in the legend for helping the readers to remember what the comparison is.

REVIEWER	Olalekan Uthman University of Warwick
REVIEW RETURNED	09-Dec-2016

GENERAL COMMENTS	1. The authors did not describe the meta-analysis methods at all in the method section. Though, this was mentioned that have been reported in the PRISMA checklist to be on page 4 • No mention was made on how they pool the studies, fixed or random-effect meta-analysis? • How was heterogeneity assessed? • Heterogeneity not also reported 2. The summary measures for the outcomes were not reported. How was the muscle strength reported, using the same scale or unit? 3. The authors stated they used Cohen kappa to assess agreement between the two authors, but the results not reported

REVIEWER	Ryan Simmons Duke University USA
REVIEW RETURNED	09-Dec-2016

GENERAL COMMENTS	Looks good overall! Just a couple of very minor quibbles from this statistician: 1) I had a difficult time connecting the results in the "Quantitative synthesis" section of the results and the various meta-analysis figures. I had to do a lot of scrolling back and forth. It would make it easier for readers if you cite a specific figure along with each reported p-value (e.g. for citing the P=0.02 for Group 2 analysis of

	the SPPB test, specifically cite Figure 6 in the text, so readers know exactly where to look, instead of making them scroll through all 14 figures to find it). 2) It wasn't clear to me exactly how the p-values were calculated in the "Qualitative synthesis" section. The figures for the quantitative analysis clearly show what statistical test was used, but neither the text nor Tables 7-9 clearly show what test was used for the qualitative analyses. I think it would be improved if there were either a quick sentence at the beginning of the qualitative synthesis section (or a footnote on Tables 7-9, which should also include the p-values). 3) According to the PRISMA checklist, the abstract should include study eligibility criteria and the systematic review registration number.

REVIEWER	Nicola Luigi Bragazzi School of Public Health, Department of Health Sciences (DISSAL), University of Genoa, Genoa, Italy
REVIEW RETURNED	25-Dec-2016

GENERAL COMMENTS	The authors have performed a systematic review and meta-analysis, according to the PRISMA guidelines. Search strategies, including the list of used keywords, the searched databases, the inclusion/exclusion criteria, are described in details in such a way to ensure reproducibility and transparency of the method. Authors have appraised the quality and the level of evidence of included studies using the GRADE and the risk of bias analyses. As I have been specifically asked to review the statistics of the paper, I have reproduced the findings of the authors and I do feel they are scientifically sound and represent a good contribution to the field, meeting with the high scientific standards of BMJ Open journal. As such, I am happy to advice acceptance of the manuscript.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1

1. “Clearly defining what efficacy comparison each of the two groups of trials tried to answer will help to quickly familiarize readers with what groups are being compared. I think group 1 is describing the additive benefit of vitamin D supplementation when combined with strength training. And I think group 2 is describing the additive benefit of strength training when combined with vitamin D supplementation. I think this specificity of phrasing in the abstract and main paper will help readers beyond what current text is used to describe the two groups of trials”.
We thank the reviewer for their positive comments regarding the relevance of this review. The clarity of the manuscript in terms of specific phrasing is a very valid point; one which has been raised by other Reviewers. Therefore, the comparator groups have been more clearly specified in the Abstract (lines 56-59) and Results sections (lines 178-181).

2. “It would be helpful if the authors could confirm that all of the studies in group 1 used a placebo control. In table 2, the exercise and vitamin D protocols are described, but the alternative control conditions are not, and this may have implications for interpreting the efficacy comparisons”.

We agree with the Reviewer that if intervention and control conditions were tabulated more clearly, interpretation of the results would be easier for the reader. Therefore, Table 2 has been modified to reflect this.

3. “My biggest concern with trials in the area is that many outcomes are measured and they are selectively reported. For example, DXA imaging can provide dozens of variables about muscle, fat, and bone. Could the authors comment on what was the pre-specified primary outcome declared on clinicaltrials.gov for each trial and was the primary endpoint was met? Did other outcomes happen to sneak into papers that weren’t declared a priori as secondary or supportive outcomes? I worry that a lot of what is being reported may be false-positive signals, therefore discussing what were planned primary and secondary outcomes versus what was ultimately was reported will be of incredible value to the field”

We thank the reviewer for highlighting this important point. The protocol for this review (registered on the PROSPERO website reference number CRD42015020157) included muscle strength as the primary outcome measure. Secondary outcomes listed were musculoskeletal function (e.g. SPPB, TUG), muscle power, body composition and bone density, serum vitamin D and calcium status; all of which were reported.
For the studies included within this review, selective reporting was addressed as part of the risk of bias analysis. All studies were assigned either a “low risk” or “unclear risk” risk of bias judgement; although all studies reported the listed outcomes, studies without a pre-defined protocol registered on clinicaltrials.gov (or similar) were judged as having an “unclear risk” of selective reporting bias.

4. “I would strongly discourage the authors from discussing or highlighting changes that were significant within only one group. For example, I think, in table 3 there are several outcomes with significant p values but make mention of “no between-group difference”. The authors have the opportunity to highlight that significant within-group changes are often presented to salvage a group-by-time interaction that was not statistically significant in an RCT (see: Allison 2016 Obesity: Common Scientific and Statistical Errors in Obesity Research, as an example). I would limit the column in table 3 labeled significant results to only those results where a significant group by time effect was demonstrated as this is the unbiased efficacy comparison derived from an RCT”.

We thank the Reviewer for raising this very important point, and for suggesting the very informative Allison et al. paper. After reading the paper, we have amended Table 3 to reflect between-group changes, and/or have highlighted where there were no between-group differences as they were reported.

5. “The authors mention that publication bias was not detected, but no formal statistical examination of publication bias was conducted, and it appears the search was limited to only [presumably] published “journal studies.” It would be helpful if the authors could elaborate how this statement is supported, as it is very likely that null trials may not be published. If so, the lack of searching for non-published trials may mean that your efficacy estimates are biased away from the null and overestimate the effects of these interventions”.

We understand the Reviewer’s comment regarding publication bias. No formal statistical analyses were performed to detect publication bias; funnel plots are listed in the Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al., 2011) as a method to detect publication bias. However, when including less than 10 studies, funnel plots are unreliable in detecting publication bias (Sedgwick P., 2013). The Cochrane Handbook for Systematic Reviews of Interventions further suggests that the inclusion of unpublished studies may introduce additional bias, as these studies have not been strengthened by the peer-review process and may be of lower methodological quality (Higgins et al., 2011). We did try to choose our wording carefully when discussing publication bias,

stating that it was “undetected” rather than not present. However, we have amended the manuscript to reflect that although it was not possible to rule out publication bias, it was not possible to statistically analyse the effect of the bias (lines 270-276).

6. “The present study did demonstrate that select outcomes including lower-extremity muscle strength and other functional outcomes did improve in both groups of efficacy studies. If at all possible, I would encourage the authors comment on the clinical utility of what these changes mean for patients and providers. For example, is a 1 kg increase in muscle strength important for a sarcopenic older adult? This may be a challenging question to answer, which has implications for future studies in this area”.

We thank the Reviewer for this comment, and agree that the clinical utility of outcomes has important implications for clinicians and patients. This is a difficult question to answer definitively; if we were to address clinical relevance of, for example, muscle strength, the literature indicates that muscle strength declines by approximately 15% per decade. The exercise and vitamin D group in study ref. 32 demonstrated an increase in right quadriceps strength of 16.7%, thus this must be viewed as clinically significant. Similarly, study ref. 31 reported an increase in isometric strength of 14.96% after 16 weeks.
However, the results of this review are difficult to interpret in this context, and significant clinical change may be obscured by variations in measurement (Roberts et al., 2011) which was often the case for the studies included within this review particularly with respect to muscle strength/size.

7. “I think the authors should provide recommendations as to what work needs to be done to definitively answer the lingering questions in this area. What study designs are useful (i.e., 2x2 factorial)? What endpoints are the most persuasive to patients and providers? Now that the authors have summarized the current data, they should provide guidance as to what they feel is the best approach moving forward. On a related note, the field of pharmaco-therapy for cancer cachexia is going through similar growing pains (see: Fearon 2015 J Cachexia Sarcopenia Muscle) and probably many lessons from cachexia with respect to study design can be applied to older adults in this setting”.

We thank the Reviewer for highlighting this important point. We have now added our suggestions for future research in more detail within the Conclusion (lines 503-516).

Reviewer #2

1. “They remark that there is lack of reviews evaluating the effects of vitamin D and exercise, but there is also an obvious reason for this; there are no trials to be included. Of note, there is one review/meta-analysis of Minshull et al (Calcif Tissue Int, 2016) which should be mentioned in this paper”.

We thank the Reviewer for this comment and for suggesting the Minshull et al paper which reviewed the role of vitamin D on neuromuscular remodelling following exercise and injury. Regarding the point that there are no studies eligible for inclusion, we now cover this within the Conclusion (lines 453-456). The Minshull paper is discussed within the Discussion section of the review, as they also comment about the lack of eligible studies for their review, the high levels of heterogeneity between studies due to differing methodologies and suggest that there is a need for additional high-quality studies within this area (lines 491-494).

2. “The most important is that the paper should be easier to read. There are so many groupings and so many test that it is difficult to follow when you are speaking about exercise effects, and when vitamin D effects. Please, pay attention to the clarity of the text. The groups are clearly stated in the

beginning of the Discussion, but elsewhere it is not easy to follow the text".

We thank the Reviewer for this comment; the point of readability was also raised by reviewer #1. Please see the response to Reviewer #1 comment number 1 above.

3. "The trials included in this meta-analysis are divided into two groups, and in most outcomes there are only two studies included. Group 1 compares vitamin D vs. placebo among exercisers. Group 2 compares exercise vs. non-exercise among participants who all were vitamin D supplemented. Neither of the groups included a pure control group with placebo and no exercise. This means that placebo or learning effects cannot be ruled out. Actually, there are only two studies which could answer to the above mentioned question (refs. 21 and 32)".

We thank the Reviewer for this valid comment; this is now clarified within the Conclusion section regarding future recommendations (lines 503-516).

4. "It would be very helpful for the readers, if you could describe the methodology a little bit more. You have most likely followed the Cochrane instructions, but a short review of the methods could be useful. For instance, how was taken into account the different study sizes, number of participants ranging from 20 to 409, and the duration of the intervention from 12 weeks to 24 months? Which variables influence on the weight percent used in the analysis?"

We thank the Reviewer for this comment; we have followed Cochrane Handbook guidelines throughout this review. We agree with Reviewer #2 that a more descriptive methodology would be helpful for the reader (Reviewer #3 also made this point). Therefore, we have added this information to the manuscript (lines 300-309).

5. "The title is not accurate. The analyses compare effects of vitamin D vs. placebo among exercisers, or exercise effects among vitamin D users, but no combined effects".

We thank the Reviewer for this comment. We have carefully considered the title and have decided to keep as is, as we believe it does reflect the subject of the review. However, we agree with a related comment that, except for 2 studies, there was no 'true' control and have responded to this separately (see comment 3).

6. "Introduction is confusing. You have referred to several papers relating to sarcopenia, which is not the study question. You should concentrate in physical functioning. A normal age-related change in physical functioning or muscle strength is not sarcopenia".

We thank the Reviewer for this comment. We have considered this comment and maintain that the referral to several papers on sarcopenia is justified and relevant to the study question. We aimed to investigate the effect of combined resistance exercise training and vitamin D3 supplementation on muscle strength and muscle function in older adults, and as such, the Introduction cites a number of papers stating their importance in terms of terms of counteracting the consequences of sarcopenia (defined as age related loss of muscle size and strength). We view that the association between maintenance/improvement of function and combating sarcopenia is implicit.

7. The conclusion is too strong. Based on this review it is not possible to speak about additive effect.

We thank the Reviewer for this comment. We do agree that based on the results of the review, it is not possible to make firm conclusions regarding the additive effects of combined vitamin D3 supplementation and resistance exercise training; therefore, we use the phrasing "tentative", and suggest that further work needs to be completed in order to reach a definitive answer.

8. “Why the paper of Kukuljan et al. is not included? They had a 2x2 factorial design with fortified milk with vitamin D and calcium, and exercise. It would be more useful for comparison with Agergaard et al, because these are the only studies including only men. Sex is probably more confounding factor than protein intake, at least in this case because protein intake was sufficient in all groups”.

We respect the Reviewer’s comment. In the pre-registered protocol (Available from http://www.crd.york.ac.uk/PROSPERO /display_record.asp?ID=CRD42015020157) we stated a priori that we would not include studies utilising nutritional supplements containing protein or other anabolic agents as we would not be able to rule out any anabolic effects resulting from the supplement and not from the resistance exercise in combination with vitamin D3. The inclusion of this particular study would mean that other studies using nutritional supplements would be eligible for inclusion, and we feel that in this instance, the results would be more difficult to interpret. The point about sex being a greater confounder that protein intake is interesting; there are a number of studies which have reported sex differences in older adults with respect to basal muscle protein synthetic rate and responsiveness to anabolic stimuli (Smith G et al., Med Sci Sports Exerc. 2012 July ; 44(7): 1259–1266 , Biol Sex Differ. 2012 May 23;3(1):11) but this was not the purpose of our review. However, in response to Reviewer #1 comment 7 to recommend future research, we do suggest analysing data for men and women separately or including sex as a covariate.

9. “Ref. 36 (Verschueren et al.) is not included in any of the analysis. Why not?”

We thank the Reviewer for this comment. This interesting study is actually included within the qualitative analysis (see reference 36).

10. “Table 2 shows number of participants in each study, but the given numbers are those who completed the study, not the randomized participants. However, the trials have used intention-to-treat analyses. Include also the sex of the participants in the table. Please, open all abbreviations in the footnotes”.

Thank you for bringing this to our attention. Table 2 has now been amended to reflect these comments.

11. “Table 3: Please add the reference numbers after the name of the author and year.”

Thank you for bringing this to our attention. Table 3 has now been amended to reflect these comments.

12. “Page 12 and Summary (page 13): Please, be careful with the directions of the changes. At least Uusi-Rasi et al. (ref. 21) did not find an improved normal walking speed with vitamin D. On the contrary, walking speed deteriorated in all groups except the group with exercise without vitamin D. The same was true with 5 x chair stand time, which improved more in the exercise groups, and no change with vitamin D without exercise was found”.

We apologise for the error in interpreting these results, and thank the Reviewer for bringing this to our attention. We have now rectified this mistake in the manuscript (lines 346-347; 446-447).

13. “Of note, in addition to ref 21, references 32 and 33 have used SPPB-test, which includes normal walking speed and 5 x chair stand test. The results should be available if requested”.

We thank the Reviewer for their suggestion. We did contact Dr Bunout for additional data, and he very kindly supplied us with a complete raw data file. Unfortunately, the SPPB test total score was

reported, rather than the individual components of the test.

14. “Bunout et al. (ref. 32) did not show an improvement in TUG-time in any of the groups. TUG-test time increased in all groups indicating that it took longer to do the test. Ref. 21 also showed declined TUG time in non-exercisers with vitamin D, and no change in exercisers with vitamin D”.

Again, we apologise for the error and thank the Reviewer for bringing this to our attention. We have now rectified this mistake in the manuscript (Table 3).

15. “Bunout et al (ref. 32) showed actually a greater increase in quadriceps strength in the exercisers without vitamin D than in the exercisers with vitamin D. In the study ref. 21 the increase was similar in exercisers with and without vitamin D. A small study of Agergaard et al. (ref. 31) cannot have such a strong effect on results. That is why it is difficult to agree with the authors that “A greater increase of muscle strength in replete older adults represents a novel finding of this review”. Compared to what? (Page 13/14). These results only confirm the previous finding that exercise improves muscle strength. Maybe the best conclusion is that these studies were unsuitable for comparison (page 14). It is difficult to include these three studies in the same analysis, because all studies have used different methods”.

We thank the Reviewer for bringing this error to our attention; Table 3 has now been amended. The results of the meta-analysis for group 1 muscle strength of the lower limb show a significant benefit for the intervention group. Study ref 31 had the smallest weighting of the 3 included studies (7.1% vs a combined weighting of 92.9% for studies 21 and 32), thus had the smallest effect on the results for this particular outcome. As the 3 studies utilised different methodologies for measuring this outcome, to attempt to account for this, a standardized mean difference rather than mean difference was calculated. However, the result of this outcome was considered within the context of the GRADE and risk of bias analyses, as serious inconsistency and moderate heterogeneity suggested that these studies may not have been suitable to compare (discussed within lines 415-421).

16. “How did you calculate the mean (sd) for the muscle strength? They are not presented as such in the papers”.

Data for muscle strength are all represented in Newton metres (Nm); data from studies 21 and 32 have been converted to Nm. Additional data were requested and kindly supplied by Dr Agergaard.

17. “Discussion for group 2 should include studies showing exercise effects (training vs. no training) on musculoskeletal health, because in this group all participants were supplemented with vitamin D. Now there is discussion about vitamin D effects, all studies (44-48) evaluated vitamin D, not exercise. The difference between exercisers vs. non-exercisers can be a learning effect, not vitamin D effect. Even if a vitamin D effect existed, this design could not show it, because all participants were supplemented”.

We thank the Reviewer for this comment. We feel that this point is now addressed within the Conclusion (lines 501-516).

18. “Page 15, narrative analysis: Too much attention is given to a muscle quality in narrative analysis, the only study using this outcome was the pilot study of Agergaard et al. (ref. 31), in which the number of older men was 7 vs. 10”.

We thank the Reviewer for this comment. We found the results for muscle quality in this paper very interesting; however, we agree that too much attention has been placed on this outcome. Therefore, we have amended the manuscript to put less emphasis on this point (lines 460-464).

19. “Page 9, 1st and 2nd paragraphs: There is also an inaccuracy in hip BMD measurements. Most studies have reported femoral neck BMD, not hip BMD. See also figure 4 and 13”.

Thank you for bringing this error to our attention. All references to hip BMD in text and Tables/Figures have now been amended.

20. "In summary: Please conclude the results separately for Group 1 and Group 2, preferably in different sections, to make it easier for readers to get the message”.

We thank the Reviewer for this comment. Results for Groups 1 and 2 are currently reported separately.

21. “Table 4: Please, add the reference numbers in the column Author, year”.

We thank the Reviewer for this suggestion. We have now amended Table 4 accordingly.

22. “Table 5: Please, use the same reference numbers of the studies throughout the paper.”

Thank you for bringing this error to our attention. This has now been rectified.

23. “Figures 2-5, and figures 6-14: It is not obvious where the numbers come for the mean (D) of the intervention or control groups. Could you please clarify? All studies do not give the exact end point values but the change from baseline to the end (adjusted change of percent). Have you calculated the means or requested the data from the authors? Keep also the lay-out similar in all tests; in all other figures the control is on the left and Intervention on the right, but TUG is shown vice versa”.

Values reported are mean percentage changes. These values were either reported in the included studies or additional information was requested from the authors. The Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al., 2011) suggests that use of change values removes a component of between-person variability from the analysis. Additionally stating, “there is no statistical reason why studies with change-from-baseline outcomes should not be combined in a meta-analysis with studies with final measurement outcomes when using the (unstandardized) mean difference method in RevMan”.
The layout of the meta-analysis is presented as such to indicate which direction would demonstrate a beneficial effect for the intervention group (i.e., intervention group on the right, an increase score will be beneficial for the intervention group).

24. “Table 7 and 8: Please give the description of the group 1 and group 2 in the legend for helping the readers to remember what the comparison is.”

Thank you for the suggestion; this information has now been added to Tables 7-9.

Reviewer #3

1. “The authors did not describe the meta-analysis methods at all in the method section. Though, this was mentioned that have been reported in the PRISMA checklist to be on page 4”.

Thank you for this comment. This information has now been added to the methods section (line 300-309)

2. “No mention was made on how they pool the studies, fixed or random-effect meta-analysis?”

Thank you for this comment. This information has now been added (line 302)

3. “ How was heterogeneity assessed?”

This is now mentioned on line 307.

4. “Heterogeneity not also reported”

Heterogeneity is already reported in Tables 5 and 6 and also within the meta-analyses (Figures 2-14).

5. “The summary measures for the outcomes were not reported. How was the muscle strength reported, using the same scale or unit?”

Thank you for this comment. This information has now been added (line 301-302)

6. “The authors stated they used Cohen kappa to assess agreement between the two authors, but the results not reported”

We thank the Reviewer for bringing this to our attention; we have now reported the inter-rater reliability (line 153).

Reviewer #4

We would like to thank Reviewer #4 for their generous comments.

1. “I had a difficult time connecting the results in the "Quantitative synthesis" section of the results and the various meta-analysis figures. I had to do a lot of scrolling back and forth. It would make it easier for readers if you cite a specific figure along with each reported p-value (e.g. for citing the P=0.02 for Group 2 analysis of the SPPB test, specifically cite Figure 6 in the text, so readers know exactly where to look, instead of making them scroll through all 14 figures to find it)”.

Thank you for this comment. We have now amended the “Qualitative synthesis” section as suggested to ease the burden on the reader (lines 317-321).

2. “It wasn't clear to me exactly how the p-values were calculated in the "Qualitative synthesis" section. The figures for the quantitative analysis clearly show what statistical test was used, but neither the text nor Tables 7-9 clearly show what test was used for the qualitative analyses. I think it would be improved if there were either a quick sentence at the beginning of the qualitative synthesis section (or a footnote on Tables 7-9, which should also include the p-values)”.

Within the qualitative synthesis, p-values stated in text are those cited within the published papers to which they relate. The qualitative synthesis is an adapted narrative review, which aims to describe and summarise the main outcomes from the included studies which were not suitable for inclusion within the meta-analyses. We would like to iterate that no additional statistical analyses have been performed on the data included within these outcomes.

3. “According to the PRISMA checklist, the abstract should include study eligibility criteria and the systematic review registration number”.

Thank you for bringing these points to our attention; this information has now been added to the Abstract (lines 45-50).

Reviewer #5

The authors have performed a systematic review and meta-analysis, according to the PRISMA guidelines. Search strategies, including the list of used keywords, the searched databases, the inclusion/exclusion criteria, are described in details in such a way to ensure reproducibility and transparency of the method. Authors have appraised the quality and the level of evidence of included studies using the GRADE and the risk of bias analyses. As I have been specifically asked to review the statistics of the paper, I have reproduced the findings of the authors and I do feel they are scientifically sound and represent a good contribution to the field, meeting with the high scientific standards of BMJ Open journal. As such, I am happy to advice acceptance of the manuscript.

We thank Reviewer #5 for their generous remarks and would like to thank them for their time spent reviewing the statistical aspects of this manuscript.

References

Cruz-Jentoft, A. J., Landi, F., Schneider, S. M., Zúñiga, C., Arai, H., Boirie, Y.,& Sieber, C. (2014). Prevalence of and interventions for sarcopenia in ageing adults: a systematic review. Report of the International Sarcopenia Initiative (EWGSOP and IWGS). Age and ageing, 43(6), 748-759.

Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]: The Cochrane Collarboration, 2011.

Roberts, H. C., Denison, H. J., Martin, H. J., Patel, H. P., Syddall, H., Cooper, C., & Sayer, A. A. (2011). A review of the measurement of grip strength in clinical and epidemiological studies: towards a standardised approach. Age and ageing, afr051.

Sedgwick, P. (2013). Meta-analyses: how to read a funnel plot. Br Med J, 346, f1342.

VERSION 2 – REVIEW

REVIEWER	Justin C. Brown Dana-Farber Cancer Institute, Boston, MA, USA
REVIEW RETURNED	13-Feb-2017

GENERAL COMMENTS	The authors have carefully attended to the comments provided on the first submission. The authors have also thoroughly revised the manuscript to address the comments and concerns from the other reviewers. Consequently, the strength of this manuscript is substantively improved from the initial submission.

REVIEWER	Kirsti Uusi-Rasi The UKK Institute for Health Promotion Research, Finland
REVIEW RETURNED	17-Feb-2017

GENERAL COMMENTS	The authors have adequately addressed the issues raised by the reviewers. However, there are a couple of errors the authors need to correct. The authors give the same effect size for both group 1 and group 2 in the text (page 15) for muscle strength of the lower limbs (2.69;

	0.95 to 4.42). Figures 2 and 8 are correct. In page 16, the authors write that “Within study [21], normal walking speed and the 5-time chair stand deteriorated non-significantly in both groups. The changes are shown correct in table 8, but the interpretation in the text is not right. Walking speed declined slightly in both vitamin D treated groups, with and without exercise. When walking speed declines (slow down), it really means deterioration. However, when chair-time declines, it means improvement.

REVIEWER	Ryan Simmons Duke University, Durham, NC, USA
REVIEW RETURNED	03-Mar-2017

GENERAL COMMENTS	Looks good!

VERSION 2 – AUTHOR RESPONSE

We would like to take the opportunity to thank the Reviewers both for their time and for their very valuable insights into this review; we are grateful to them and confident that, as a result of their comments, the quality of the manuscript has been strengthened.
We have responded to each of the Reviewer’s comments below; any amendments are shown in tracked-changes within the manuscript, and where appropriate, line numbers indicate the location of resulting changes.

BMJ Open Editorial Team:

1. The in text citation for Reference 37 is missing on your main document file. Please amend accordingly.
Thank you for bringing this to our attention. Reference 37 has now been removed. As such, the total number of references in the manuscript has been reduced from 54 to 53.

2. Please upload each Figure file separately.
These files have now been uploaded.

1. The authors give the same effect size for both group 1 and group 2 in the text (page 15) for muscle strength of the lower limbs (2.69; 0.95 to 4.42). Figures 2 and 8 are correct.

We thank the reviewer for bringing this error to our attention. This has now been corrected (page 15, line 317).

2. In page 16, the authors write that “Within study [21], normal walking speed and the 5-time chair stand deteriorated non-significantly in both groups. The changes are shown correct in table 8, but the interpretation in the text is not right. Walking speed declined slightly in both vitamin D treated groups, with and without exercise. When walking speed declines (slow down), it really means deterioration. However, when chair-time declines, it means improvement.

We thank the reviewer for this comment. We have now amended this sentence (page 16, lines 345-346).