PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	How Does a Physical Activity Program in Elementary School Affect Fracture Risk? A Prospective Controlled Intervention Study in Malmo, Sweden
AUTHORS	Coster, Marcus; Fritz, Jesper; Nilsson, Jan-Åke; Karlsson, Caroline; Rosengren, Bjorn E.; Dencker, Magnus; Karlsson, Magnus

VERSION 1 - REVIEW

REVIEWER	Jon Minton University of Glasgow, UK
REVIEW RETURNED	23-May-2016

GENERAL COMMENTS	I think this is an interesting and well written paper addressing an important question about the role of PA in reducing fracture risk. I have a series of more minor comments but my major concerns are as follows: 1) Presentation of results: Three schools comprise the control group. Although it is stated the schools are neighbouring and similar socioeconomically, no data is presented which allows the socioeconomic characteristics of the control schools to be compared. i) I recommend publishing details on socieconomic characteristics of schools in an appendix. Further, results from all three control schools are pooled together making it impossible to know how similar or different rates in the three control schools are from each other. ii) I recommend the results be presented separately for each school in an appendix. iii) Additionally, a simple graph showing the IRR by year for control group(s) and intervention groups is recommended. This last point leads me onto the second area of concern: 2) Interpretation of results it appears from table 1 that fracture rate stars off highest in year 1 in the intervention group, then falls to around 20-22/1000 person years, though raising to 34.7 in the 6th year. iv) I would recommend the 6th year difference in intervention be discussed further. By contrast the trend in the control group is increasing. The numbers themselves show that fracture rates in the control group are lower than in the intervention group for the first four years, statistically significantly so in the 1st year. This more complex set of findings isn't signaled by the generally pro-intervention wording in the abstract and elsewhere. I think a more even-handed interpretation of the results would be more equivocal of the results, especially where the results show they are associated with a statistically significant increase of 65% in fracture rate in the 1st year. iv) I recommend redrafting the text throughout to make the more complex and equivocal findings more apparent, especially within the

	abstract. In terms of theory/literature: v) I would recommend more discussion of the two following topics: 1) generalisability of findings to other countries and marginal effects of further PA: it seems even the control group may be more physically active than many schools in other countries, and this should be addressed. Are the effects of going from very low to some PA marginally more effective than from some to much PA (See a BMJ Open paper of mine where I suggest yes). 2) Differences between different types of PA. I would imagine some types of PA will increase risk of fracture due to contact, others may be low impact but not lead to increased bone strength so not prevent it either, and yet more (the 'ideal' school PA) may both reduce fracture risk in the short term without raising it in the short term. Further discussion about different types of PA, and the type of PA taught in these schools and this intervention, are required to be able to claim benefits of school PA effectively.

REVIEWER	Harri Sievanen, Research Director The UKK Institute for Health Promotion Research Tampere, Finland
REVIEW RETURNED	09-Jun-2016

GENERAL COMMENTS	In this large long-term non-randomized physical activity (PA) intervention trial, all children who started the first grade in four schools were recruited in the study during years 1998 -2012. Altogether, 3535 children took part in this trial. The active group (N=1339, one school) underwent 40 min moderate PA during each school day while the control group (N=2195, three schools) followed the standard physical education curriculum 60 min per week. Fractures and their causes were extracted from reliable health register source. Incidence rate ratio between the groups showed a declining trend (based on somewhat dubious correlation analysis – see comment #6) and reached statistical significance by the 8th year of intervention suggesting that about a half of fractures (IRR = 0.48) may be prevented by regular physical training. This is a highly important finding and fully consistent with what is known mainly from older adults (eg the meta-analysis by El-Khoury et al BMJ 2014) It is a bit unfortunate that the fractures occurring specifically during the training program could not be identified. This would have allowed one to assess the relative harms, costs and benefits of daily training. Apparently, this was probably not a real issue. In addition, the lack of relevant information, including habitual PA and dietary intake of essential nutrients (Ca, vitamin D, protein) are essential limitations and potential confounders of the study as the authors also admit. A crucial question is whether potential between-group differences in these variables would have affected the present observations. 1. Page 3, line 45, What was the rationale of choosing the references #1-4 as representative studies of relevant interventions? Rather, adequately powered RCTs (eg Action Schools BC! by McDonald et al in JBMR 2007) should be cited and given as examples. 2. Page 3 ,line 50. I wonder whether all the references 8-10 are relevant in terms of actual childhood PA and fracture risk later in life. Recent studies by Tolonen S et al (Bone 2015) and Nordstrom et al (JBMR 2013) are relevant in this respect and should be cited as well.

	3. Page 5, line 34. The text indicates that some children began the 1st grade and started the intervention in 2012. If so, the intervention period of these late starters could not have been eight years. Perhaps I have misunderstood the text. In any case, please clarify and rephrase as appropriate. This applies also to the abstract and title. 4. Page 5, line 45. What does the expression “this event” refer to? Please specify. 5. Page 6, line 32. The correlation does not solely apply to the intervention group, since IRR is calculated from fracture data of both groups. Thus thee beginning of the sentence may be omitted. 6. I doubt whether the correlation between IRR and the year of intervention is relevant here. Only the data from the two last years indicate a sudden change in IRR – the first six do not show even clear indication of such a trend. The possibility that the observation is a result of a chance may not be ruled out. I also wonder what is the explanation for the large variation in N values in each year. I would have expected that the number of participants with an 8-year follow-up would have been larger because the study started already in 1998 and many participants would have been followed all 8 years that there would not have been so much loss in N in each year. As stated by the authors (page 3, line 14), there were no dropouts, which is amazing. Please explain. It would be interesting the see the fracture incidence in groups with complete data from all eight follow-up years (ie, the N would be virtually the same for each year). 7. Data should be analyzed separately for boys and girls, because there is evidence that the responses to childhood PA and exercise can be different (see McDonald et al JBMR 2007 and Tolonen et al Bone 2015) 8. Table 2. Is there any possibility to estimate the costs of fractures in both groups? It appears that the distribution of fractures is not quite similar. 9. Page 9, line 41. Is there any evidence for higher bone mass or better neuromuscular function during the 8-year follow-up? These data were apparently collected and may be presented as appropriate to support the arguments.

REVIEWER	Hang Lee, PhD Massachusetts General Hospital Biostatistics Center and Harvard Medical School, Boston, MA, USA
REVIEW RETURNED	19-Jul-2016

GENERAL COMMENTS	Statistical comment: The overall trend of eight-year change in IRR could have been better characterized if a more sophisticated statistical method had been applied. It is notable that except Year 6, the time longitudinal patterns of the incidence rates in both groups appeared to be consistent with the study hypothesis. Fitting a random effects Binomial or Poisson (over-dispersed Poisson model may work better) regression model to the person time data could have provided very sensible estimates of within group time dependencies of the incidence rates over time as well as a statistically more powerful overall between group comparison of the longitudinal IRRs. Please consider performing such an analysis by using either of the aforementioned models to regress the annual incidence on year (1, 2, 3, ..., 8), PA (1 if yes; 0 if no), year × PA, and year 6 (1 if year 6; 0 otherwise), which would be instrumental to improve your articulation

	of the study findings.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Jon Minton
Institution and Country: University of Glasgow, UK
Competing Interests: None declared

I think this is an interesting and well written paper addressing an important question about the role of PA in reducing fracture risk.
I have a series of more minor comments but my major concerns are as follows:

Comment 4: Presentation of results: Three schools comprise the control group. Although it is stated the schools are neighbouring and similar socioeconomically, no data is presented which allows the socioeconomic characteristics of the control schools to be compared.
i) I recommend publishing details on socieconomic characteristics of schools in an appendix.
Answer: We have in previous publications from the POP-study reported background data, including socioeconomical and ethnical status, in a sub-cohort of the cohort used in this paper (Valdimarsson, 2006; Lindén 2006). This sub-cohort consisted of children with school start during the first three years of the study (1998 – 2000). Since this manuscript is based on a larger cohort and is a register study, we do not have any background data for all children included in the study. We believe that the children in the large cohort would reflect the sub-cohort, but since we cannot confirm this we have decided to exclude this comment from the manuscript. The sentence now is a follows, please see page 4, lines 23-25. “The schools were community-based, government-funded, used the compulsory standard national curriculum, and were located within the same city region. The children were assigned to their school according to their residential address.”

Further, results from all three control schools are pooled together making it impossible to know how similar or different rates in the three control schools are from each other.
ii) I recommend the results be presented separately for each school in an appendix.
Answer: As recommended by the reviewer we have performed analyses per control school and have included the results in the appendix, please see Appendix 4. These analyses suggest that the trend for incremental fracture incidence increase by school year. Acceleration during the final years is apparent in each of the control schools. During the eighth year in the Ribbersborg school the fracture incidence however did not seem to increase as much as in the other control schools.

iii) Additionally, a simple graph showing the IRR by year for control group(s) and intervention groups is recommended.
Answer: As suggested we have included a graph of IRR by school year, please see Figure 1, page 10.

This last point leads me onto the second area of concern:
Comment 5: Interpretation of results
it appears from table 1 that fracture rate stars off highest in year 1 in the intervention group, then falls to around 20-22/1000 person years, though raising to 34.7 in the 6th year.
i) I would recommend the 6th year difference in intervention be discussed further.
Answer: As recognized by the reviewer the fracture incidence in the intervention group seems to be higher during the sixth year of intervention than the previous and following years. We have found no explanation for this finding. The intervention was not changed and we could not see a trend in types of fractures or trauma mechanism. The finding is most likely a result of chance, but factors such as maturation, growth, and lifestyle changes may have contributed. The following sentences are included

in the discussion section to highlight the sixth year fracture rate, please see page 10, lines 20-24.
“During the sixth year of the study the fracture incidence in the intervention group seemed to increase compared to previous and following years. We have found no explanation for this finding and it may most likely be a result of chance, but we cannot exclude that factors such as maturation, growth, and lifestyle changes have influenced the results.”

By contrast the trend in the control group is increasing. The numbers themselves show that fracture rates in the control group are lower than in the intervention group for the first four years, statistically significantly so in the 1st year. This more complex set of findings isn't signaled by the generally pro-intervention wording in the abstract and elsewhere. I think a more even-handed interpretation of the results would be more equivocal of the results, especially where the results show they are associated with a statistically significant increase of 65% in fracture rate in the 1st year.
ii) I recommend redrafting the text throughout to make the more complex and equivocal findings more apparent, especially within the abstract.
Answer: We have redrafted the manuscript and included a wider perspective of the findings from the 1st year of the study, please see abstract, page 2.
“Results: During the 1st year after initiation of the intervention the fracture IRR was 1.65 (1.05, 2.08) (mean 95 % CI). For each year of the study the fracture incidence rate in the control group compared to the intervention group increased by 15.7 % (5.6, 26.8 %) (mean 95 % CI). After 8 years the IRR of fractures was 52 % lower in the intervention group than in the control group [IRR 0.48 (0.25, 0.91) (mean 95 % CI)]. Conclusions: Introduction of the school-based intervention program is associated with higher fracture risk in the intervention group during the 1st year followed by a gradual reduction so that during the 8th year the fracture risk was lower in the intervention group.”
and page 10, lines 2-5, “We have previously reported an inverse correlation between number of years of PA intervention and fracture risk after a temporary increase in fracture risk when introducing the PA intervention,(3) but we can now also state that the annual fracture risk is 52% lower during the eighth year of intervention.”
and page 10, lines 9-13, “There was no difference in overall fracture risk during the study period between the intervention group and control group. However, as previously reported, (3) year-by-year fracture risk evaluation revealed an initial transient increase in fracture risk in the intervention group followed by a gradual fracture risk reduction leading to a significantly lower fracture risk at the end of the intervention period.”
and page 13, lines 1-4. “In conclusion, this is the first time it has been shown that a childhood intervention with extra PA during a longer period is associated with a lower fracture risk. We must however emphasize that this is preceded by a transient higher fracture risk after introduction of the PA.”

In terms of theory/literature: I would recommend more discussion of the two following topics:
Comment 6: generalizability of findings to other countries and marginal effects of further PA: it seems even the control group may be more physically active than many schools in other countries, and this should be addressed. Are the effects of going from very low to some PA marginally more effective than from some to much PA (See a BMJ Open paper of mine where I suggest yes).
Answer: The control group received 60 minutes of physical education per school week throughout the school year (37 weeks). We must emphasize that our conclusions are based on comparisons with this group of children and we cannot state whether or not the same effects will be seen in comparison to other groups with either lower or higher amounts of PA. To clarify this we have included a statement in the limitations section, please see page 12, lines 18-22. “Furthermore, we must emphasize that our conclusions are based on comparisons with a group of children receiving 60 minutes of physical education per school day and with fairly generous amount of spare time activity. Thus, we cannot state whether or not the same effects will be seen in comparison to other groups with either lower or higher amounts of baseline PA.”

We agree with the reviewer in that the same amount of extra activity in inactive children probably will have greater effects than in already active children. Whether the intervention would be a fixed amount of extra PA (as in this study) or an increase to a certain PA level effects would be dependent on the baseline activity level as well as the spare-time activity level during the study period, which dictates the relative increase. It is also important to highlight that this PA program was designed to facilitate the participation of all children irrespective of previous sport experience or activity level. The idea was to get all children to a minimum activity throughout the study, but also to hopefully promote an interest in PA with more activity in spare-time.

Comment 7: Differences between different types of PA. I would imagine some types of PA will increase risk of fracture due to contact, others may be low impact but not lead to increased bone strength so not prevent it either, and yet more (the 'ideal' school PA) may both reduce fracture risk in the short term without raising it in the short term. Further discussion about different types of PA, and the type of PA taught in these schools and this intervention, are required to be able to claim benefits of school PA effectively.
Answer: The physical activity classes taught in the schools followed the national curriculum and consisted of various ordinary school PE activities. These activities included various ball games (including football, handball, basketball and more), track and field activities, running, jumping, swimming, and various games (for example catch the flag). The activities were not designed specifically for bone or muscle strengthening, but rather so that they could easily be implemented in every school. One of the reasons for including a variety of activities was that we believe that it is important that the children do not get bored during the PE classes by the activities and then do not participate at full strength.

As the reviewer states different physical activities have different effects on bone strength and this is probably also true for fracture risk. Other types of activities than those in our intervention program may have a better fracture preventive effect and may not influence the initial fracture risk as much. Now that we have established a positive effect by our program future studies must try to elucidate if other PA programs have even greater effects and still enable all children to participate, are fun for the children, and have long term fracture protective effects perhaps without increasing the short term fracture risk.

For clarification the following sentences are included in the manuscript, please see page 5, lines 6-9, “The PE was supervised by the regular schoolteachers and consisted of ordinary PE activities within the national educational plan such as running, jumping, climbing, swimming, and various ball games. Thus, the PE was not specifically designed to reduce fractures.”
and page 13-14, lines 23-7. “Furthermore, as described in the methods section the PA intervention consisted of regular physical activities not specifically designed to reduce fractures. An intervention with more high-impact activities, which strengthen bone and muscle, may result in an even lower fracture risk, but could also result in higher exposure to trauma and thereby a higher fracture risk. Also, there is a risk that repeated activities make the children bored and thereby reduce their intensity levels during the PE. Greater effects may have also been possible if the PA was changed during the course of the program. All in all, the optimal program would include all children, prevent fractures throughout the study and not result in a higher fracture risk after initiation. To find out which activities and what levels of intensity that have the best fracture protective effects further studies are needed.”

Reviewer: 2
Reviewer Name: Harri Sievanen, Research Director
Institution and Country: The UKK Institute for Health Promotion Research, Tampere, Finland
Competing Interests: None declared

In this large long-term non-randomized physical activity (PA) intervention trial, all children who started the first grade in four schools were recruited in the study during years 1998 -2012. Altogether, 3535 children took part in this trial. The active group (N=1339, one school) underwent 40 min moderate PA during each school day while the control group (N=2195, three schools) followed the standard physical education curriculum 60 min per week. Fractures and their causes were extracted from reliable health register source. Incidence rate ratio between the groups showed a declining trend (based on somewhat dubious correlation analysis – see comment #6) and reached statistical significance by the 8th year of intervention suggesting that about a half of fractures (IRR = 0.48) may be prevented by regular physical training. This is a highly important finding and fully consistent with what is known mainly from older adults (eg the meta-analysis by El-Khoury et al BMJ 2014)
It is a bit unfortunate that the fractures occurring specifically during the training program could not be identified. This would have allowed one to assess the relative harms, costs and benefits of daily training. Apparently, this was probably not a real issue. In addition, the lack of relevant information, including habitual PA and dietary intake of essential nutrients (Ca, vitamin D, protein) are essential limitations and potential confounders of the study as the authors also admit. A crucial question is whether potential between-group differences in these variables would have affected the present observations.

Comment 8: Page 3, line 45, What was the rationale of choosing the references #1-4 as representative studies of relevant interventions? Rather, adequately powered RCTs (eg Action Schools BC! by McDonald et al in JBMR 2007) should be cited and given as examples.
Answer: The rationale was to include references of long-term physical activity intervention studies in children. Preferably we would want to refer to long term RCTs, but we have been unable to find any such studies. The suggested article suits the reasoning very well and has been added as a reference in the manuscript.

Comment 9: Page 3, line 50. I wonder whether all the references 8-10 are relevant in terms of actual childhood PA and fracture risk later in life. Recent studies by Tolonen S et al (Bone 2015) and Nordstrom et al (JBMR 2013) are relevant in this respect and should be cited as well.
Answer: As suggested we have updated the references regarding adolescent PA and fracture risk later in life. We have added the articles above as references and also removed a reference that was found less adequate than the suggested articles.

Comment 10: Page 5, line 34. The text indicates that some children began the 1st grade and started the intervention in 2012. If so, the intervention period of these late starters could not have been eight years. Perhaps I have misunderstood the text. In any case, please clarify and rephrase as appropriate. This applies also to the abstract and title.
Answer: The inclusion of children was ongoing from year 1998 until 2012. So, as the reviewer indicates, only the children starting school between years 1998 – 2005 have finished eight years of intervention. Children starting from year 2006 and later were included in the study for as many years that they had completed when the study period ended in august 2013. To clarify this in the manuscript we have added the following sentence in the methods section, please see page 5, lines 17-19.“Consequently, the children were followed from in-between one and eight years depending on which year they began 1st grade. The number of participants having completed a year of intervention is thus lower by each successive school year (Table 1).”

Comment 11: Page 5, line 45. What does the expression “this event” refer to? Please specify.
Answer: This event refers to if children moved out of the region or changed schools between intervention and control schools and thereby ended their follow-up in the study. We have now clarified this with the sentence below, please see page 5, lines 19-21. “Children who moved out of the region or changed school between intervention and control schools (n=183) were followed until the date either of these things occurred.”

Comment 12: Page 6, line 32. The correlation does not solely apply to the intervention group, since IRR is calculated from fracture data of both groups. Thus thee beginning of the sentence may be omitted.
Answer: We agree with the reviewer that this expression applies to both groups and we have now removed the initial part of the sentence, please see page 6, lines 20-21.

Comment 13:
i) I doubt whether the correlation between IRR and the year of intervention is relevant here. Only the data from the two last years indicate a sudden change in IRR – the first six do not show even clear indication of such a trend. The possibility that the observation is a result of a chance may not be ruled out.
Answer: The correlation analysis was added to show that after the initial increase in IRR during the first school year there was a correlation between an additional school year and a lower IRR compared to the previous year (although not compared to before the intervention until the final years). So although only the data from the last two years indicate a lower IRR there is a correlation throughout the study period as a result from the higher IRR during the first school year. We believe that this statistical analysis gives a measure of uncertainty and thus adds important knowledge and gives us the possibility to present the uncertainty for the results in two different ways.

Finally, as the reviewer suggests there is always a possibility that statistical significance is the result of chance, most often accepted as a level of 5 %. During the eighth year there was a statistically significant lower IRR inferring that there is an actual difference between groups.

ii) I also wonder what is the explanation for the large variation in N values in each year. I would have expected that the number of participants with an 8-year follow-up would have been larger because the study started already in 1998 and many participants would have been followed all 8 years that there would not have been so much loss in N in each year.
Answer: The inclusion of children into the study was ongoing from year 1998 until 2012. So, as the reviewer indicates, only the children starting school between years 1998 – 2005 have finished eight years of intervention. Children starting from year 2006 and later were included in the study for as many years as they had completed when the study period ended in august 2013. So for every year in the study there is a loss of one annual school cohort of children, thus explaining the drop in N each year. Also, please see answer to comment 10.

iii) As stated by the authors (page 3, line 14), there were no dropouts, which is amazing. Please explain.
Answer: The physical activity intervention program was conducted in an elementary school. In the country the study was conducted it is compulsory for all children to attend school classes including the PE classes. Consequently, all children had to attend the physical activity program throughout the study. Therefore, unless children moved from the region they continued the intervention program for all eight years. Furthermore, since we included children at different periods of time we could also include the children who moved out of the region until that moving date. Thus, formally there were no dropouts. Finally, as we collected information regarding fractures through radiographic archives using social security numbers we could collect information of all participating children in both the intervention school and the control schools. Thus, no children were asked to report any fractures enabling us to include all children.

iv) It would be interesting the see the fracture incidence in groups with complete data from all eight follow-up years (ie, the N would be virtually the same for each year).
Answer: We agree with the reviewer that the ideal would be to use only children who completed the full eight years. However, due to the fact that the number of fractures is few during the pre-pubertal

years compared to the pubertal years such an inclusion results in low power with large confidence intervals, mostly during the initial years of the study. We therefore decided to include all available children. However, for comparison we have now included a table of only those children participating for the full eight years as an appendix, please see Appendix 1.

Comment 14: Data should be analyzed separately for boys and girls, because there is evidence that the responses to childhood PA and exercise can be different (see McDonald et al JBMR 2007 and Tolonen et al Bone 2015)
Answer: We agree with the reviewer that gender specific analyses would be preferable. As the reviewer indicates there might be different responses to physical activity between the sexes and also maturation is different for boys and girls during this period. However, unfortunately we are concerned that we do not have enough power for gender specific analyses. Regardless of the high risk of type II errors we have now at the recommendation of the reviewer also made gender specific analysis and included result tables for boys and girls separately in the appendix, please see Appendix 2 and 3. We leave the decision on whether or not to include these tables in the manuscript to the editor and the reviewer.

Comment 15: Table 2. Is there any possibility to estimate the costs of fractures in both groups? It appears that the distribution of fractures is not quite similar.
Answer: As noticed by the reviewer there seems to be a slight difference in distribution of fractures between groups. We are however unable to decide whether or not this is a coincidence or something associated with the intervention program. A determination of the costs of the fractures in the different groups would be most interesting. However, since the majority of these fractures do not require costly treatments such as surgery or admission to hospital, and most not even a check-up, the overwhelming part of the costs will emanate from the emergency department visit. These costs are difficult to estimate and may even be similar for many fracture types. Even though interesting this investigation and discussion are not within the scope of this manuscript.

Comment 16: Page 9, line 41. Is there any evidence for higher bone mass or better neuromuscular function during the 8-year follow-up? These data were apparently collected and may be presented as appropriate to support the arguments.
Answer: Since reviewers for previous articles from the POP-study were concerned about the different cohorts for fracture assessment and musculoskeletal findings (these analyses were performed in a sub-cohort of about 250 individuals) they recommended us to not report bone mass and muscle function data together with the fracture data. The concern does have merit and after this recommendation we decided not to include the data in this manuscript.

Reviewer: 3 (stats reviewer)
Reviewer Name: Hang Lee, PhD
Institution and Country: Massachusetts General Hospital Biostatistics Center and Harvard Medical School, Boston, MA, USA
Competing Interests: None declared

Statistical comment:

Comment 17: The overall trend of eight-year change in IRR could have been better characterized if a more sophisticated statistical method had been applied. It is notable that except Year 6, the time longitudinal patterns of the incidence rates in both groups appeared to be consistent with the study hypothesis. Fitting a random effects Binomial or Poisson (over-dispersed Poisson model may work better) regression model to the person time data could have provided very sensible estimates of within group time dependencies of the incidence rates over time as well as a statistically more

powerful overall between group comparison of the longitudinal IRRs. Please consider performing such an analysis by using either of the aforementioned models to regress the annual incidence on year (1, 2, 3, ..., 8), PA (1 if yes; 0 if no), year × PA, and year 6 (1 if year 6; 0 otherwise), which would be instrumental to improve your articulation of the study findings.
Answer: As suggested by the stats reviewer we have conducted a new statistical analysis to better characterize the overall trend of IRR change during the study. We used a Generalized Estimating Equations (GEE) model with exchangeable working correlations for this purpose. This analysis estimated that after initiation of the study the fracture incidence rate in the control group was 52.9 % (95 % CI (35.1, 79.7 %)) of the fracture incidence rate in the intervention group.
Furthermore, during the study period the fracture incidence rate for the control group compared to the intervention group increases each year of the study by 15.7 % (95 % CI (5.6, 26.8 %). Since the reviewer had some concerns regarding the higher IRR during the sixth year we have also conducted this analysis using a dummy variable for the sixth year, which results in a 0.1 percentage point difference (15.8 % (95 % CI (5.7, 26.9 %))).

We have now included the following sentences in the manuscript, please see the statistical section page 6, lines 11-15,"We also used a Generalized Estimating Equations (GEE) model with exchangeable working correlations to estimate the difference in fracture incidence rates during the overall study period. Finally, as there was an outlier in fracture incidence in the intervention group during the sixth school year we conducted the same analysis using a dummy variable for this outlier for comparison." and the results section page 6-7, lines 24-5. "Using the GEE model we estimated that after initiation of the study the fracture incidence rate in the control group was 52.9 % (95 % CI (35.1, 79.7 %)) of the fracture incidence rate in the intervention group. Furthermore, during the study period the fracture incidence rate for the control group compared to the intervention group increased each year of the study by 15.7 % (95 % CI (5.6, 26.8 %)). Using a dummy variable in the analysis resulted in a 0.1 percentage point difference in the estimation (15.8 % (95 % CI (5.7, 26.9 %)))."

With these changes we hope that the manuscript could be considered for publication in BMJ Open.