PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Exposure to heavy physical work from early to later adulthood and primary health care visits due to musculoskeletal diseases in midlife: a register linked study
AUTHORS	Halonen, Jaana; Shiri, Rahman; Mänty, Minna; Sumanen, Hilla; Solovieva, Svetlana; Viikari-Juntura, Eira; Kahonen, Mika; Lehtimaki, Terho; Raitakari, Olli; Lallukka, Tea

VERSION 1 – REVIEW

REVIEWER	Esben Meulengracht Flachs Department of Environmental and Occupational Medicine, Bispebjerg University Hospital, Copenhagen, Denmark
REVIEW RETURNED	21-May-2019

GENERAL COMMENTS	I have some comments to the authors to clarify various areas of the text: Abstract: No comments. Introduction: Precise and succinct description of research field and aim. Methods: Selection of participants clearly described, however 1170 participants is mentioned twice, both as those answering both times on work load questions, and as those with no missing on any covariates. Please clarify if these numbers should be equal or not. A related point is on table 1, where I would prefer to have a table with information on actual participants (n=1170), rather than on the entire study population, or even on both populations to be able to identify possible differences. I would omit sex specific information, as that is not used anywhere else. Covariates: I guess that physical activity relates only to leisure time physical activity, as occupational activity is the exposure, please clarify. Some of the covariates eg. BMI and smoking are mentioned as time-varying, though the latest information is collected at start of follow-up (2011), and thus cannot (?) vary in the follow-up period from 2011-2014. Some information on loss to follow-up 2011-2014 is needed. The methods section describes the 1170 participants as those with complete covariate information, whereas the covariates section

	mentions imputation of missing covariates, please clarify which statement is correct. If imputation of missing data has been carried out I would like more information on how an imputation of eg. mean level of parents education or ever/never smoker is used in the analysis. It seems to me that eg. an average of the categories ever/never smoker is meaningless. If imputation of missing data is used my advice would be to use more modern methods (i.e. multiple imputation), instead of just using mean levels. Statistical analysis: It is not clear from the description which analysis has been carried out. The text mentions GEE poisson regression, but neither the outcome nor statistics section mentions repeated measurements. Rather the outcome section mentions time at first diagnosis as end of follow up, which implies no repeated events. The statistics section then mentions cox regression as alternatives, but the outcome section do not mention time-to-event as an outcome, only diagnoses. I would like the statistics section to be thoroughly rewritten and the outcome section to comply with this. Poisson regresson with offset to account for time at risk or cox-regression with time to event as outcome seems like good choices for statistical analysis. Risk time should be counted only from start of follow up in 2011. Results: This section is well-written. Discussion: In general a well written and fair discussion of methods and results in relation to other findings. I miss some comments on: Missing data and potential impact on results. Healthy worker selection particularly in relation to the inclusion criteria which demands employment at both early an late adulthood. This means that persons who have left employment - maybe because of musculoskeletal disease or pain - are left out of analyses. As work load is only determined at two points, all the potential effect of work load between these points is not addressed in the study. Impact of work load or employment status during follow up from 2011 to 2014. Conclusion: The conclusion seems like a fair assessment of the results and discussion.

REVIEWER	Daniel Holman University of Sheffield, United Kingdom.
REVIEW RETURNED	05-Jun-2019

GENERAL COMMENTS	Review of exposure to heavy physical work form early to later adulthood and primary care visits due to musculoskeletal diseases in midlife Overall this seems like a useful study on an important topic, but I have some methodological concerns, as well as other various comments. I believe the paper requires quite a bit of work in light of these, but it might be of publishable standard if this work was to be undertaken. My specific comments are:

	P4.L8. I'm not sure it is true that MSD are the leading cause of work disability/disability retirements, as mental health problems/depression are often shown to be on par. See example the report https://www-gov-uk.sheffield.idm.oclc.org/government/publications/working-for-a-healthier-tomorrow-work-and-health-in-britain or more widely the recent WHO report https://www.who.int/news-room/detail/30-03-2017--depression-let-s-talk-says-who-as-depression-tops-list-of-causes-of-ill-health or the recent paper https://oem.bmj.com/content/early/2019/04/15/oemed-2018-105493.abstract. P6.L24. Dichotomising the question about physical work seems to represent quite a loss of information, especially given that this is the main exposure in the analysis. E.g. I would expect a big difference between medium heavy work and very heavy work. This strategy needs justifying, and perhaps testing to see the implications of throwing away this amount of information. Is it possible to run an analysis using all information to see how this influences the results? P7.L47. I'm curious why parental occupational status in childhood and not own occupational status was used as a covariate. Wouldn't own occupational status be more relevant as a confounder? If the reason is that it explains why people have different physical workloads and thus reduces the association to null, them a similar mechanism will be operating with respect to parental occupational status, which significant influences own occupational status. P7.L50. Similarly with health behaviours e.g. smoking, physical activity, as well as BMI. I don't quite understand the rationale for adjusting for these. Generally confounders should be associated with both the exposure and the outcome, but the association between health behaviours, physical work, and healthcare visits is unclear. It may well be important to adjust for, but again more justification is needed. P8.L10. 'Plenty of missing values' is vague P8.L13. Imputing 'using the mean of the study sample' is a weak imputation strategy and generally not advised. Why was multiple imputation not used? P9.L43. It would be good to see the differences with the Cox model, perhaps in an online appendix. However, it is important to note that GEE and Cox models involve different interpretations, assumptions, and pitfalls. Thus the choice to use GEE should be justified – currently no justification or literature is discussed. P10.L10. MSDs were not objectively measured in this study – they are measured by primary care visits. Whether people seek health care and whether they are diagnosed with MSD depends on a whole range of factors, some of which are subjective. This needs to be acknowledged in the paper more widely, in terms of the implications for the analysis, results and interpretation. P11.L38. 'neither' should be 'not' P11.L43. 'already' should be deleted. P11.L49. Some authors have only contributed materials, tools, or funding. To me that doesn't justify authorship. The statistician who helped form the dataset probably has more of a claim to authorship. P16. P5 says 1170 participants but this table lists 5850. A flow chart would really help to understand how the sample was selected. P17. The top right section of the table is wrong as the % for osteoarthritis for all do not add to 100. This table includes 1120 participants – why the discrepancy with 1170? Further the abstract states 1061. Again, flow diagram needed.

	P19. Is this Figure 1? It is not labelled. Further it should be replaced with a flow diagram – it currently has no numbers or inclusion/exclusion criteria. P20. The authors should be commended for including a STROBE statement. However I found the reference to page number ranges slightly frustrating as it required me to read 2-3 pages of text to see where the criterion was evidenced. Also, item 3 is not on p5 (hypotheses) – at least not clearly stated, and 12d seems applicable to me. 16c needs justifying. The whole table to be carefully checked – I may have missed other oversights.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Esben Meulengracht Flachs Institution and Country: Department of Environmental and Occupational Medicine, Bispebjerg University Hospital, Copenhagen, Denmark Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below I have some comments to the authors to clarify various areas of the text:

1. Abstract: No comments.

2. Introduction: Precise and succinct description of research field and aim.
OUR RESPONSE: Thank you for this positive feedback.

Methods:
3. Selection of participants clearly described, however 1170 participants is mentioned twice, both as those answering both times on work load questions, and as those with no missing on any covariates. Please clarify if these numbers should be equal or not.
OUR RESPONSE: We have now removed the part of sentence referring to follow-up time on page 5. We have also corrected the value for the total study sample to include only those with data on all covariates, see pages 5-6: "After excluding those with missing data on any covariate (after re-coding and imputation), the final study sample was 1056 cohort participants (with 5171 observations) who all had work exposure measurement from early adulthood and at least one measurement from later adulthood."

We have also replaced Figure 1 with a flow chart of the sample selection and re-ran models 1 in Table 3 with the complete case study sample (i.e., N individuals= 1056).

4. A related point is on table 1, where I would prefer to have a table with information on actual participants (n=1170), rather than on the entire study population, or even on both populations to be able to identify possible differences. I would omit sex specific information, as that is not used anywhere else.
OUR RESPONSE: Thank you for the suggestion. We have revised Table 1 so that it describes the data at the individual level.

5. Covariates: I guess that physical activity relates only to leisure time physical activity, as occupational activity is the exposure, please clarify.
OUR RESPONSE: We have clarified this as suggested on page 8: "Measure for leisure-time physical activity (PA) was based on a set of questions requesting the frequency and intensity of PA, frequency of vigorous PA, hours spent on vigorous PA, average duration of a PA session, and participation in organized PA."

6. Some of the covariates eg. BMI and smoking are mentioned as time-varying, though the latest information is collected at start of follow-up (2011), and thus cannot (?) vary in the follow-up period from 2011-2014.
OUR RESPONSE: This is true. We had no information on any of the covariates after 2011 but smoking and BMI could vary between the baseline and 2011 along with the exposure.

We have now mentioned the possible consequences of this in the Discussion on page 12: The follow-up period for the outcomes was not very long and unobserved changes in the exposure or covariates during the outcome follow-up could have caused some bias to the findings resulting in under- or overestimation of the observed associations."

7. Some information on loss to follow-up 2011-2014 is needed.
OUR RESPONSE: As we used register data, the follow-up for the outcomes is rather complete. Only if a person emigrates, she/he will no more have registered health care visits. In Discussion on page 12, we have written: "Moreover, the used cohort data were representative of the general population with relatively little loss to follow-up.15 As we used register data, only persons who emigrate will no more have registered health care visits."

8. The methods section describes the 1170 participants as those with complete covariate information, whereas the covariates section mentions imputation of missing covariates, please clarify which statement is correct.
OUR RESPONSE: We have now corrected this section on page 5: "After excluding those with missing data on any covariate (after re-coding and imputation), the final study sample was 1056 cohort participants (with 5171 observations) who all had work exposure measurement from early adulthood and at least one measurement from later adulthood."

We have also revised the part describing covariates on page 8: "For PA, we used the maximum of the three measurements of the PA index in adulthood (2001, 2007 and 2011), as these data had plenty of missing values (N missing =730 in 2001, 150 in 2007 and 475 in 2011), but the patterns of PA have been observed to remain constant in adulthood.18 Missing data on smoking were re-coded as "non-smoker", and missing data on BMI were imputed using mean of the study sample in the corresponding survey."

Correspondingly, we corrected the Model 2 adjustments on page 9: "Model 2 was additionally adjusted for parental occupational class, physical activity, and time-varying smoking and BMI."

9. If imputation of missing data has been carried out I would like more information on how an imputation of eg. mean level of parents education or ever/never smoker is used in the analysis. It seems to me that eg. an average of the categories ever/never smoker is meaningless. If imputation of missing data is used my advice would be to use more modern methods (i.e. multiple imputation), instead of just using mean levels.
OUR RESPONSE: We have now revised this part in the text on page 8 as follows: "Missing data on smoking were coded as "non-smoker", and missing data on BMI were imputed using mean of the study sample in the corresponding survey. Although this is not the strongest imputation method we considered it the most applicable one, as the method only concerned one covariate."

While using mean for imputation is not the strongest method for imputation, we only used this for one variable that was used as a covariate. Other methods are neither without limitations, for example, multiple imputation includes an assumption that data are missing at random, which rarely is true. We did not impute data for parental occupational status or physical activity.

Statistical analysis:
10. It is not clear from the description which analysis has been carried out. The text mentions GEE poisson regression, but neither the outcome nor statistics section mentions repeated measurements. Rather the outcome section mentions time at first diagnosis as end of follow up, which implies no repeated events. The statistics section then mentions cox regression as alternatives, but the outcome section do not mention time-to-event as an outcome, only diagnoses. I would like the statistics section to be thoroughly rewritten and the outcome section to comply with this.
OUR RESPONSE: We apologize for the ambiguity. In the paragraph describing outcomes we have now clarified the outcomes used, please see pages 6-7: "We examined primary health care visits due to a musculoskeletal diagnosis. The follow-up started from the day after returning wave 3 survey in 2011. Repeated visits were used for the main analyses. For an alternative analysis, time to the first visit was used as an outcome, and the follow-up from returning the survey in 2011 continued until the first primary health care visit, death (from Statistics Finland) or end of the follow-up (end of 2014), whichever occurred first."

On page 8 we have clarified the use of GEE model: "We used generalized estimating equation (GEE) models with Poisson distribution to assess associations between the five-class physical work exposure, "no exposure" serving as the reference group, and repeated primary health care visits due to MSD."

And on page 9 the outcome used for the alternative analyses: "As an alternative method, we ran the analyses using Cox proportional models using time to the first visit as the outcome, which resulted in very similar findings (supplemental Table 1)."

11. Poisson regression with offset to account for time at risk or cox-regression with time to event as outcome seems like good choices for statistical analysis. Risk time should be counted only from start of follow up in 2011.
OUR RESPONSE: This is true and the follow-up for the outcomes started after returning the survey in 2011, please see page 6: "The follow-up started from the day after returning wave 3 survey in 2011."

Results:
12. This section is well-written.
OUR RESPONSE: Thank you for the positive feedback.

Discussion:
In general a well written and fair discussion of methods and results in relation to other findings. I miss some comments on:
13. Missing data and potential impact on results.
OUR RESPONSE: Missing data for physical work exposure may have caused the healthy worker effect, which is now discussed on page 12: "Some healthy worker effect may have attenuated the findings as we required minimum of two responses (from early and later adulthood) regarding physical heaviness of work and those with physically strenuous work or with musculoskeletal problems may have left employment before the second survey."

14. Healthy worker selection particularly in relation to the inclusion criteria which demands employment at both early and late adulthood. This means that persons who have left employment - maybe because of musculoskeletal disease or pain - are left out of analyses.
OUR RESPONSE: Please see the response above.

15. As work load is only determined at two points, all the potential effect of work load between these points is not addressed in the study. Impact of work load or employment status during follow up from 2011 to 2014.

OUR RESPONSE: It is true that there is possibility for the exposure (as well as outcomes) to have changed between the surveys, and the exposure may have changed also during the outcome follow-up. We have mentioned this in the Discussion on page 12: “We cannot rule out the possibility of changes in the exposure or outcomes between the survey waves, which may have caused under- or over-estimation of the associations.”
AND: “The follow-up period for the outcomes was not very long and unobserved changes in the exposure or covariates during the outcome follow-up could have caused some bias to the findings resulting in under- or overestimation of the observed associations.”

Conclusion:
16. The conclusion seems like a fair assessment of the results and discussion.
OUR RESPONSE: Thank you for the positive feedback.

Reviewer: 2
Reviewer Name: Daniel Holman
Institution and Country: University of Sheffield, United Kingdom.
Please state any competing interests or state ‘None declared’: None declared.

Overall this seems like a useful study on an important topic, but I have some methodological concerns, as well as other various comments. I believe the paper requires quite a bit of work in light of these, but it might be of publishable standard if this work was to be undertaken. My specific comments are:

1. P4.L8. I’m not sure it is true that MSD are the leading cause of work disability/disability retirements, as mental health problems/depression are often shown to be on par. See example the report https://www-govuk.sheffield.idm.oclc.org/government/publications/working-for-a-healthier-tomorrow-work-and-health-in-britain or more widely the recent WHO report https://www.who.int/news-room/detail/30-03-2017--depression-let-s-talk-says-who-as-depression-tops-list-of-causes-of-ill-health or the recent paper https://oem.bmj.com/content/early/2019/04/15/oemed-2018-105493.abstract.
OUR RESPONSE: It is true that mental health problems are increasingly prevalent. Still, in Finland at least, musculoskeletal disorders were the leading cause of disability retirement in 2018 (see ref 3). The WHO news indicated by the Reviewer includes total population statistics and does not distinguish between employees who are at risk of work disability and those who might not be entitled, e.g., to sickness absence benefits. However, we have now revised the text on page 4 so that mental health problems are also mentioned: “Musculoskeletal diseases (MSD) are, along with mental disorders, the leading cause of work disability1 measured as sickness absence2 and disability retirements.3”

2. P6.L24. Dichotomising the question about physical work seems to represent quite a loss of information, especially given that this is the main exposure in the analysis. E.g. I would expect a big difference between medium heavy work and very heavy work. This strategy needs justifying, and perhaps testing to see the implications of throwing away this amount of information. Is it possible to run an analysis using all information to see how this influences the results?
OUR RESPONSE: We agree it would be interesting and important to include as much information as possible, and to study also gradient in the associations, including light, medium heavy and heavy work separately. However, as we examined repeated exposure to physical work using 2-3 measurements, it was most practical to use the dichotomized variable for grouping the exposure (no, only early, only later, inconsistent, early and later). The exposure categories for the repeated exposure measure would become very complex if more categories were used (e.g. 3*3 for light, medium and heavy). In addition, the proportion of those with heavy/very heavy physical work at

baseline was 9.8%, thus, categorizing the exposure using this cut-off might have resulted in low power.

We have now mentioned the loss of information and its possible effects on the findings in the Discussion on pages 11-12: “We also used a dichotomized physical heaviness of work measure where medium and heavy/very heavy work were combined as the proportion of those with heavy/very heavy work was rather low (10% at early adulthood). The used cut-off may have attenuated the observed associations if medium heavy work had substantially weaker association with health care visits than heavy/very heavy work.”

3. P7.L47. I’m curious why parental occupational status in childhood and not own occupational status was used as a covariate. Wouldn’t own occupational status be more relevant as a confounder? If the reason is that it explains why people have different physical workloads and thus reduces the association to null, them a similar mechanism will be operating with respect to parental occupational status, which significant influences own occupational status.
OUR RESPONSE: Parental occupational status was used as an indicator for childhood socioeconomic status that may be associated with the occupational exposures in adulthood as well as with poor health. We have revised the text regarding possible links between parental occupational status and health and work exposures in the introduction on pages 4-5: “…prior studies have rarely accounted for family background although parental socioeconomic position has been linked, for example, to later musculoskeletal problems11 and widespread pain12 as well as career possibilities13 and choices.14”

As own occupational status (or education) is highly correlated with physical workload, we could not adjust for own occupational status.

4. P7.L50. Similarly with health behaviours e.g. smoking, physical activity, as well as BMI. I don’t quite understand the rationale for adjusting for these. Generally confounders should be associated with both the exposure and the outcome, but the association between health behaviours, physical work, and healthcare visits is unclear. It may well be important to adjust for, but again more justification is needed.
OUR RESPONSE: Smoking, physical activity and BMI are all known risk factors of musculoskeletal disorders. Smoking can be also considered as an indicator of low socioeconomic position, which may affect the choice of employment and further, physical workload. This has now been clarified in the Methods section on page 8: “All these covariates have been linked to back problems in prior studies11 19-21 and smoking can be also considered as an indicator of low socioeconomic position, which may affect the choice of employment and further physical workload.”

However, as it is true that all the included health-related variables may not be true confounders, we show both sex and age adjusted effect estimates as well as those including the other covariate adjustments.

5. P8.L10. ‘Plenty of missing values’ is vague
OUR RESPONSE: We have now clarified the numbers of missing values for PA in each survey, please see page 8: “as these data had plenty of missing values (N missing =730 in 2001, 150 in 2007 and 475 in 2011),…”

6. P8.L13. Imputing ‘using the mean of the study sample’ is a weak imputation strategy and generally not advised. Why was multiple imputation not used?
OUR RESPONSE: We have now revised this part in the text on page 8 as follows: “Missing data on smoking were re-coded as “non-smoker”, and missing data on BMI were imputed using mean of the

study sample in the corresponding survey. Although this is not the strongest imputation method we considered it was the most applicable one, as the method only concerned one covariate."

While using mean for imputation might not be the strongest/best method for imputation, we only used this for one variable that was used as a covariate. Other methods are neither without limitations, for example, multiple imputation includes an assumption that data are missing at random, which rarely is true.

7. P9.L43. It would be good to see the differences with the Cox model, perhaps in an online appendix. However, it is important to note that GEE and Cox models involve different interpretations, assumptions, and pitfalls. Thus the choice to use GEE should be justified – currently no justification or literature is discussed.
OUR RESPONSE: On page 8 we have now added justification for the chosen method: "This method was chosen as the GEE models permit specification of a working correlation matrix that accounts for the form of within-subject correlation of responses on dependent variables of many different distributions, including Poisson.20"

Results from the Cox proportional hazards model have been added to a supplemental material. This has now been mentioned on page 9: "As an alternative method, we ran the analyses using Cox proportional hazard models using time to the first visit as the outcome, which resulted in very similar findings (supplemental Table 1)."

8. P10.L10. MSDs were not objectively measured in this study – they are measured by primary care visits. Whether people seek health care and whether they are diagnosed with MSD depends on a whole range of factors, some of which are subjective. This needs to be acknowledged in the paper more widely, in terms of the implications for the analysis, results and interpretation.
OUR RESPONSE: We have now elaborated the discussion on page 12 considering the possible subjective nature of the outcome measure: "It can be speculated that primary health care visits with musculoskeletal diagnosis in midlife are mostly a result of pain complaints. Severe pain may interfere with work activities and induce need for sickness absence, which may be the primary motivation for the visit to a physician. Thus, the used outcomes may reflect the severity of work disability due to a subjective measure of musculoskeletal pain."

9. P11.L38. 'neither' should be 'not'
OUR RESPONSE: We have replaced neither with not as suggested.

10. P11.L43. 'already' should be deleted.
OUR RESPONSE: We have deleted this word as suggested.

11. P11.L49. Some authors have only contributed materials, tools, or funding. To me that doesn't justify authorship. The statistician who helped form the dataset probably has more of a claim to authorship.
OUR RESPONSE: This is not quite true, please see the end of the Authors' contributions: "All authors were involved in interpretation of the findings, writing the paper and approved the submitted and published versions."

12. P16. P5 says 1170 participants but this table lists 5850. A flow chart would really help to understand how the sample was selected.
OUR RESPONSE: We apologize for the ambiguity. The 1170 referred to the number of individuals/participants in the data, whereas 5850 referred to the observations in the repeated data. We have now replaced Figure 1 with a flow chart of the sample selection.

13. P17. The top right section of the table is wrong as the % for osteoarthritis for all do not add to 100. This table includes 1120 participants – why the discrepancy with 1170? Further the abstract states
OUR RESPONSE: Thank you for pointing this out. The percentages for osteoarthritis in Table 2 were indeed incorrect. We have now updated the whole Table 2 as we checked the sample selection. The "1120" (currently 1083) refers to the number of health care visits (observations) in the repeated data and thus is different form the number of individuals (N=1056) included in the data set.

14. 1061. Again, flow diagram needed.
OUR RESPONSE: We have now replaced Figure 1 with a flow chart of the sample selection.

15. P19. Is this Figure 1? It is not labelled. Further it should be replaced with a flow diagram – it currently has no numbers or inclusion/exclusion criteria.
OUR RESPONSE: We have now replaced Figure 1 with a flow chart of the sample selection. Figure legend is given before the References section, but we have now added the legend also to the figure itself.

16. P20. The authors should be commended for including a STROBE statement. However I found the reference to page number ranges slightly frustrating as it required me to read 2-3 pages of text to see where the criterion was evidenced. Also, item 3 is not on p5 (hypotheses) – at least not clearly stated, and 12d seems applicable to me. 16c needs justifying. The whole table to be carefully checked – I may have missed other oversights.
OUR RESPONSE: We have now checked that the page numbers match between the STROBE and the revised manuscript. We have also added a sentence on page 5 clarifying the hypothesis.

VERSION 2 – REVIEW

REVIEWER	Esben Meulengracht Flachs Department of Environmental and Occupational Medicin, Bispebjerg University Hospital, Copenhagen, Denmark
REVIEW RETURNED	29-Jul-2019

GENERAL COMMENTS	The reviewer completed the checklist but made no further comments.

REVIEWER	Daniel Holman Department of Sociological Studies, University of Sheffield
REVIEW RETURNED	15-Jul-2019

GENERAL COMMENTS	The authors have done a good job of responding to the comments. My one remaining comment is that I cannot find % missing for smoking and BMI. Is this the 28 stated on the flowchart? If so, please clarify on the flow chart, and also state in the text the number of missings for these two variables. This is important given your imputation method.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2
Reviewer Name: Daniel Holman
Institution and Country: University of Sheffield, United Kingdom.
Please state any competing interests or state ‘None declared’: None declared.

The authors have done a good job of responding to the comments.

1. My one remaining comment is that I cannot find % missing for smoking and BMI. Is this the 28 stated on the flowchart? If so, please clarify on the flow chart, and also state in the text the number of missings for these two variables. This is important given your imputation method.
OUR RESPONSE: Thank you for the positive feedback. We have now clarified in the text on page 8 the range of missing observations for BMI and smoking in different study phases that were imputed: “Number of missing observations varied by phase from 5 in 2007 to 369 in 1986, some of which were excluded due to missing data on other covariates. Missing data on BMI were imputed using mean of the study sample in the corresponding survey. Number of missing observations varied from 0 in 2001 to 411 in 1989, some of which were excluded due to missing data on other covariates.”

The missing N=28 in Figure 1 refers to missing data on physical activity and parental occupational status. We have now clarified this in the revised Figure 1.