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)	Can socio-economic health differences be explained by physical activity at work and during leisure-time? Rationale and protocol of the Active Worker individual participant meta-analysis
AUTHORS	Coenen, Pieter; Huysmans, Maaike; Holtermann, Andreas; Troiano, Richard; Mork, Paul Jarle; Krokstad, Steinar; Clays, Els; Van Mechelen, Willem; van der Beek, Allard

VERSION 1 – REVIEW

REVIEWER	Satoru Kanamori Department of Preventive Medicine and Public Health, Tokyo Medical University, Tokyo, Japan.
REVIEW RETURNED	01-May-2018

GENERAL COMMENTS	This is an interesting study that could be an essential description. Findings such as these would be able to contribute to progress in this research area. Further attention to the issues below would significantly strengthen the authors' report. In this research, does leisure-time physical activity include physical activity of commuting and housework? I thought that it would be better to describe in the background. In Figure 3, a putative causal model of the physical activity paradox is shown. Leisure-time physical activity has not only biological effect but also psychological and social effect such as enjoyment and social interaction. Please consider this point.

REVIEWER	Christine Friedenreich Alberta Health Services, Calgary, Canada
REVIEW RETURNED	16-Jul-2018

GENERAL COMMENTS	BMJ Open 2018-023379 **General Comments:** The authors describe the protocol for the *Active Worker* Study which is a planned individual participant meta-analysis of 49 studies (106 publications) that had assessed the association between occupational and leisure time physical activity on cardiovascular or all-cause mortality outcomes conducted anywhere worldwide. The IPD meta-analysis has already been registered in Prospero and the authors have carefully prepared this manuscript according to the PRISMA-P statement. Furthermore, they will be using the PRISMA-IPD methods for the reporting of the

	eventual results from this analysis. The underlying hypothesis that they wish to examine is if workers who are exposed to higher levels of occupational physical activity are at increased risk of cardiovascular and all-cause mortality because they also have lower levels of leisure time physical activity and if socioeconomic differences can explain these different outcomes. This study is a welcome addition to the literature and the authors are to be congratulated for assuming such a large research endeavour. Overall, the protocol is well described and there are only a few points for consideration. **Specific Comments:** 1. The authors have chosen to focus only on cardiovascular mortality and all-cause mortality outcomes. It seems that there is a missed opportunity here by not also including cancer mortality outcomes. Some of the co-authors are also working in the field of cancer research and could bring that expertise to this group. There is now a very strong evidence base that physical activity is related to the risk of developing and surviving after cancer. Hence, some justification and rationale for the choice of outcomes should be provided in the introduction of this paper. 2. The authors have acknowledged the potential limitations of this pooling of individual level data as well as the challenges of harmonizing the data in the bullet points that are listed at the beginning of the manuscript. They have not, however, provided much information within their methods on how they will address these limitations in their planned study. Some consideration for approaches to obtaining all of the studies as well as the methods to use to deal with data harmonization should be included in this protocol. For example, there are groups of scientific experts who have specialized in data harmonization who might be able to provide assistance and/or guidance on how to harmonize data from these 49 studies. 3. Background – the authors have use the words “magic pill” to describe physical activity for the prevention of premature death. This wording conjures up the notion of a simple fix for a non-communicable disease which is not really adequate for describing the effect of physical activity on mortality outcomes. This phrase should be replaced. 4. Figure 1 – the inclusion of a conceptual model in this protocol is a strength. However, the model itself is not very clear and could use some more thought. Specifically, the two rectangular boxes at the top and bottom of the figure that include pre-existing conditions, material factors, behavioural factors, psychosocial factors and demographic factors and their relationships with the internal box need to be more clearly described in this model. Are these outside box factors to be considered as confounders, effect modifiers, mediators or something else? Some re-design of the figure to make it very clear to a reader is necessary.

	In addition, the inclusion of a some footnote to suggest how these factors will be modelled/considered in the analysis would be helpful. 5. More detail is needed on how the data harmonization will be done. 6. The *Active Worker* consortium template should be provided as an Appendix to this paper. 7. The authors are categorizing physical activity into sedentary behaviour, low, moderate and high. No information is provided on whether they will attempt to extract the exact dose of activity in terms of frequency, intensity and duration of activity so that they might be able to quantify the exposure in terms of MET-hours/week. It would be preferable to quantify the exposures more precisely if at all possible. Some discussion of this point would be valuable to add to the methods section. 8. The authors also list in Table 1 that they will include both self-report and objective measurements of physical activity. They do not explain how they will combine these two different types of physical activity data in the text or if these analyses will be done separately by source of exposure data. 9. More description of the planned analyses should be added to provide more details on how each aim will be addressed, particularly the analyses on secondary outcome measure (e.g. which secondary outcomes will be assessed? How will the modelling be done to ensure that a full examination of confounding and effect modification is done without conducting too many sub-group analyses?). 10. Table 1 – the authors have stated that they will be seeking the socioeconomic status of the parents. This variable seems somewhat odd since most studies capture the participant's socioeconomic status, not that of their parents. Perhaps there is some misunderstanding here? 11. Figure 3 – the authors provide a putative causal model to their paper. It is unclear, however, if they will be able to assess these possible causal associations with this study. More detail and development of the rationale behind this model and how it will be tested in this study would add more value to this protocol paper. 12. Authorship – the Active Worker core research group and the advisory group of international collaborators are co-authors on this paper. Only the core research group members' author contributions are provided in this paper. 13. The dates for the planned IPD analysis are not provided in this manuscript. The timelines for this project are also not provided. 14. The authors have chosen to include a table with a list of all of the excluded studies. This level of detail is generally not provided for meta-analyses/IPD analyses and could be remove from the supplementary materials. 15. The authors have not stated if they have already begun to contact the primary authors from the 49 studies and, if so, if they have experienced any difficulty obtaining the data from these authors.

	Minor points: 1. Page 9, line 10: “send” should be “sent”. 2. Page 9, line 11: “reminding” should be “reminder”. 3. Page 9, line 48: “study” should be “studies”. 4. Page 10, line 46: delete “in order” and add an “s” to “aim”. 5. Page 10, line 50: change “aim” to “aims”. 6. Page 10, line 53: add the words “or not” after “whether”. 7. Page 11, line 6: delete “in order”. 8. Page 11, line 30: change “analysis” to “analyses”. 9. Page 12, line 20: delete “in order”. 10. Figure 3 – “respons” should be “response”

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

This is an interesting study that could be an essential description. Findings such as these would be able to contribute to progress in this research area. Further attention to the points below would significantly strengthen the authors' report.

Response: We would like to thank the reviewer for the assessment and positive appraisal of our work. In general we agree with the points raised and have revised our manuscript accordingly. A point-by-point response is provided below, with the respective changes in the manuscript highlighted in yellow.

In this research, does leisure-time physical activity include physical activity of commuting and housework? I thought that it would be better to describe in the background.

Response: Thank you for this relevant comment. We have now indicated the different domains of physical activity in the introduction section, by stating (page 5):

‘To date, physical activity is considered an important and significant behaviour for the prevention of premature death from many lifestyle-related diseases[8]. The health implications of engagement in different domains of physical activity (according to common PA research: leisure time, work, household and transport[9]) are considered to be positive and alike.’

Also, we have now in the methods section specified the type of physical activity variables that we intent to request by stating (page 9):

‘Data on occupational and leisure-time physical activity (predictor variables; typically not including the domains of transport and household) … will be retrieved from each of the participating studies.’

And in table 1, by listing:

‘Other (e.g. transport and/or household) domains of physical activity’

In Figure 3, a putative causal model of the physical activity paradox is shown. Leisure-time physical activity has not only biological effect but also psychological and social effect such as enjoyment and social interaction. Please consider this point.

Response: We agree that this is an important point to consider. We have now added a statement on this in the discussion section, which now reads (page 15):

‘Also, different mental and social health implications of work and leisure time physical activities may play a role in this. For example, leisure time physical activity is often associated with positive outcomes of mental health (such as mood states like anxiety, stress and depression)[44] and social functioning (such as loneliness and social support)[45]. On the other hand, highly demanding jobs often come with mental strain which has been shown to be associated with negative health outcomes, including cardiovascular disorders[46]’

In the figure caption of Figure 3, we have now added:

‘Also, different mental and social health implications of work and leisure time physical activities may play a role in this. For example, leisure time physical activity is often associated with

positive outcomes of mental health and social functioning, while highly demanding jobs often come with mental strain which has been shown to be associated with negative health outcomes.'

Reviewer 2

General Comments:
The authors describe the protocol for the *Active Worker* Study which is a planned individual participant meta-analysis of 49 studies (106 publications) that had assessed the association between occupational and leisure time physical activity on cardiovascular or all-cause mortality outcomes conducted anywhere worldwide. The IPD meta-analysis has already been registered in Prospero and the authors have carefully prepared this manuscript according to the PRISMA-P statement. Furthermore, they will be using the PRISMA-IPD methods for the reporting of the eventual results from this analysis. The underlying hypothesis that they wish to examine is if workers who are exposed to higher levels of occupational physical activity are at increased risk of cardiovascular and all-cause mortality because they also have lower levels of leisure time physical activity and if socioeconomic differences can explain these different outcomes. This study is a welcome addition to the literature and the authors are to be congratulated for assuming such a large research endeavour. Overall, the protocol is well described and there are only a few points for consideration.

Response: We would like to thank the reviewer for the assessment and positive appraisal of our work. In general we agree with the points raised and have revised our manuscript accordingly. A point-by-point response is provided below, with the respective changes in the manuscript highlighted in yellow.

Specific Comments:
1. The authors have chosen to focus only on cardiovascular mortality and all-cause mortality outcomes. It seems that there is a missed opportunity here by not also including cancer mortality outcomes. Some of the co-authors are also working in the field of cancer research and could bring that expertise to this group. There is now a very strong evidence base that physical activity is related to the risk of developing and surviving after cancer. Hence, some justification and rationale for the choice of outcomes should be provided in the introduction of this paper.

Response: We agree that there is a growing body of evidence showing cancer-related health effects associated with physical activity. The differential health effects of occupational and leisure time physical activity, i.e. the physical activity paradox, are less well documented for cancer-related outcomes. They are, however, well documented for cardiovascular and all-cause mortality. Because of this, hypotheses for the physical activity paradox are mainly based on cardiovascular outcomes (as described in more detail in the discussion section; page 3). Our project was based on this premise. Therefore in our grant application as well as in our search for eligible studies, we have only targeted studies on cardiovascular and all-cause mortality outcomes. Considering the above, we would respectfully advocate to stick with our original plan.

2. The authors have acknowledged the potential limitations of this pooling of individual level data as well as the challenges of harmonizing the data in the bullet points that are listed at the beginning of the manuscript. They have not, however, provided much information within their methods on how they will address these limitations in their planned study. Some consideration for approaches to obtaining all of the studies as well as the methods to use to deal with data harmonization should be included in this protocol. For example, there are groups of scientific experts who have specialized in data harmonization who might be able to provide assistance and/or guidance on how to harmonize data from these 49 studies.

Response: We have now provided more detail on the harmonization of data (see our response to point 5). Regarding the approaching of potential collaborators, we have now mentioned (page 9):
'From these eligible studies, corresponding authors, principal investigators and/or researchers from the network of the consortium members will be invited to collaborate in our Active Worker Study consortium. Invitations will be sent by email. After no response, a reminder email will be sent and/or a telephone call will be made to achieve participation. Participation in the consortium includes sharing of data and ideas, and manuscript preparation. Reasons for refusal to participate will be noted.

After agreeing to collaborate, potential collaborators will be asked to fill out our data request form in order for us to obtain more information on the respective cohorts (e.g. regarding contact details of principle investigators, study design and available data). Also, researchers will be asked to sign a policy document after which they will be asked to transfer their (anonymized and encrypted) data.'

As a response to point 13 below, we have also provided more information on the timeline of our study, including on our efforts in contacting potential collaborators.

3. Background – the authors have use the words "magic pill" to describe physical activity for the prevention of premature death. This wording conjures up the notion of a simple fix for a non-communicable disease which is not really adequate for describing the effect of physical activity on mortality outcomes. This phrase should be replaced.

Response: We have rephrased this section, it now reads (page 5):
'To date, physical activity is considered an important and significant behaviour for the prevention of premature death from many lifestyle-related diseases[8].'

4. Figure 1 – the inclusion of a conceptual model in this protocol is a strength. However, the model itself is not very clear and could use some more thought. Specifically, the two rectangular boxes at the top and bottom of the figure that include pre-existing conditions, material factors, behavioural factors, psychosocial factors and demographic factors and their relationships with the internal box need to be more clearly described in this model. Are these outside box factors to be considered as confounders, effect modifiers, mediators or something else? Some re-design of the figure to make it very clear to a reader is necessary. In addition, the inclusion of a some footnote to suggest how these factors will be modelled/considered in the analysis would be helpful.

Response: We agree with the reviewer that our figure was not clear enough. We have now amended the figure, making clearer which of the variables are potential confounders (the rectangular box at the top) and which of the variables are effect modifiers (the rectangular box on the bottom). Also, we have clarified the associations that we are going to study by labelling them a, b and c in the figure (in line with the text in the methods section; page 11). In the figure caption, we have now added:
'To address aim 1, we will assess the combined association of occupational and leisure-time physical activity with the primary outcome (pathway a). For aims 2 and 3, we will additionally consider potential modifying effects. To address research aim 4, we will study the mediating effects of occupational and leisure-time physical activity on the association of socio-economic status and the primary outcome. To do so, we will provide an estimate of the relative strength of the mediation effect (axb) and we will be comparing this to the total direct effect (c'), using the product-of-coefficients test. In all these models, potential confounders will be considered.'

5. More detail is needed on how the data harmonization will be done.

Response: We have now provided more information on the harmonization procedure. This section now reads (page 9):
'Of continuous variables, Z-scores will be calculated to be able to harmonize constructs measured by different questionnaires/items. The most detailed level of data possible will be used for data harmonization (e.g. continuous variables rather than categorical ones). However, if needed categorical variables from different studies will be harmonised, classifying them into predetermined categories.'
Also, more information regarding the harmonization of PA variables has been provided (see our response to point 7 below).

6. The *Active Worker Study* consortium template should be provided as an Appendix to this paper.

Response: The template that we were talking about in our manuscript, is a dynamic document in which different definitions of the various variables of our IPD dataset are described. This document is updated every time a new study is added to our dataset. Therefore, it currently only consists of information from two studies. As such, we do not believe that providing this information is helpful to the reader. We have therefore removed references to this template from our manuscript. We do, however, intent to make this template available at a later stage.

7. The authors are categorizing physical activity into sedentary behaviour, low, moderate and high. No information is provided on whether they will attempt to extract the exact dose of activity in terms of

frequency, intensity and duration of activity so that they might be able to quantify the exposure in terms of MET-hours/week. It would be preferable to quantify the exposures more precisely if at all possible. Some discussion of this point would be valuable to add to the methods section.

Response: We have now provided more details on this, by stating (page 10):
'For physical activity variables (both occupational and leisure-time), we intent to express all data into MET-hours/week, using established scoring systems, such as for the IPAQ questionnaire[32] and for the US Census Occupational Classification System[33], and the Compendium of Physical Activities[34]. If available, such continuous data can be harmonized across studies and modelled as such, according to the below mentioned analysis plan. In an additional sensitivity analysis, also considering studies without continuous MET-based data, physical activity data will be categorised into the four categories from the physical activity continuum (i.e. sedentary behaviour, low, moderate and high), using established cut-off values[35]. The latter will be done during a consensus meeting with the Active Worker Study core group.'

8. The authors also list in Table 1 that they will include both self-report and objective measurements of physical activity. They do not explain how they will combine these two different types of physical activity data in the text or if these analyses will be done separately by source of exposure data.

Response: We indeed had the intention to obtain both self-reported and objectively measured physical activity data in the first place. However, while contacting potential collaborating researchers, we realized that as far as we are aware there is currently no data available on objectively measured physical activity for which: a) work chores can be separated from leisure-time chores, and; b) physical activity can be linked to mortality outcome data. We have therefore chosen to remove references to objectively measured physical activity from our analysis plan. In the discussion section, we have now addressed this limitation, by mentioning (page 16):
'As far as we are aware there is currently no data available on objectively measured physical activity for which: a) work chores can be distinguished from leisure-time chores, and; b) physical activity can be linked to mortality outcome data. A limitation of the current study will therefore be that all data will be based on self-reports, which may be biased[52].'

9. More description of the planned analyses should be added to provide more details on how each aim will be addressed, particularly the analyses on secondary outcome measure (e.g. which secondary outcomes will be assessed? How will the modelling be done to ensure that a full examination of confounding and effect modification is done without conducting too many subgroup analyses?).

Response: We have now provided more details as to how each of the four research aims will be addressed. Regarding the secondary outcome measures, we have now mentioned (page 13):
'If time permits, abovementioned research aims will additionally be addressed using the secondary outcome measures.'
Also we have now added more information on the procedures for addressing confounders and effect modifiers (page 6):
'For aims 2 and 3, additional interaction terms will be included, providing information regarding potential effect modifying effects. This interaction term will gain insight into whether or not the effect of occupational and leisure-time physical activity on the outcome differs across different modifier categories. Here, potential effect modifying factors such as chronic lifestyle-related diseases, hypertension or body mass index (BMI) addressing aim 2 and factors such as cardiorespiratory fitness addressing aim 3 will be considered. In case interaction terms are associated with the outcomes in a statistically significant manner ($p<0.05$), models will be stratified by that particular effect modifier.'
and (page 11)
'In all abovementioned models (for aims 1-4), relevant confounders (e.g. demographic, behavioural, psychosocial and material factors as shown in Table 1 and Figure 1) will be considered using a selection procedure in which all potential confounders will be univariately tested, after which a backward selection procedure will be conducted among variables with a statistically significant ($p<0.10$) univariate association with the primary outcome measure. Potential confounding variables with a multivariate statistically significant ($p<0.05$) association with the primary outcome, will be retained in the final models.'

10. Table 1 – the authors have stated that they will be seeking the socioeconomic status of the parents. This variable seems somewhat odd since most studies capture the participant's socioeconomic status, not that of their parents. Perhaps there is some misunderstanding here?

Response: We agree with the reviewer and have therefore removed this variable from Table 1 as well as from Figure 1.

11. Figure 3 – the authors provide a putative causal model to their paper. It is unclear, however, if they will be able to assess these possible causal associations with this study. More detail and development of the rationale behind this model and how it will be tested in this study would add more value to this protocol paper.

Response: We have now added in more details to clarify the development of the model. We now mention (page 14):

'This figure, which is modified from earlier work[19] and is theorised from existing evidence, shows the differential health consequences of occupational physical activity (being more common among those from lower socio-economic status groups) and leisure-time physical activity (being more common among those from higher socio-economic status groups). In this model it is hypothesized[19] that …'

Moreover, we have now clarified that some, but not all parts of this model will be tested in the current study. We now mention (page 15):

'The current study will further strengthen the evidence-base regarding the health implications of occupational and leisure-time physical activity. However, mechanisms underlying this association, such as the ones shown in Figure 3, should also be investigated in studies with other research methodologies, for example in field measurements using ambulatory measurements of physical activity, biomechanical load, heart rate and blood pressure.'

12. Authorship – the Active Worker core research group and the advisory group of international collaborators are co-authors on this paper. Only the core research group members' author contributions are provided in this paper.

Response: The authors of the current manuscript include all members of the core research group and international advisory group. These are researchers who were involved in the design of the current research plan as well as in the acquisition of research funding. At the time of submission of this manuscript, no other collaborators had yet committed to our consortium, nor were they involved in the design of our IPD. These collaborators will, however, when committing to the consortium with their data and expertise, be asked to be co-author on any future publications on the outcomes of our consortium. To make this more apparent, we have now added the word 'future' to the section specifying authorship (page 12).

13. The dates for the planned IPD analysis are not provided in this manuscript. The timelines for this project are also not provided.

Response: More information on the timelines of this project has now been provided in the *'project and data management'* section. This now reads (page 13):

'The data inclusion procedure has started in May 2018 and we intend to continue doing so until the end of September 2018. As of July 20th 2018, researchers of a total of nine studies have already agreed to be part of the consortium, with two of them having transferred their data already. Of the other studies, application procedures for obtaining the data have started or studies were excluded from our consortium. Reasons for exclusion were the inability to get in contact with researchers (in some cases researchers have retired or have passed away) or the researchers not willing or being able to participate. Merging of the datasets and analysis syntaxes will be prepared in the period July-September 2018, to ensure that analyses can start when all data are collected in September 2018. Reporting on the data (in accordance to below mentioned dissemination plan) is expected to be done by Mid-2019.'

14. The authors have chosen to include a table with a list of all of the excluded studies. This level of detail is generally not provided for meta-analyses/IPD analyses and could be remove from the supplementary materials.

Response: We agree with the reviewers that this level of detail is not required. However, as this information is available and may be helpful for some readers (in an earlier published review we were asked by a reviewer to provide such a table), we would prefer to retain this table in the manuscript.

15. The authors have not stated if they have already begun to contact the primary authors from the 49 studies and, if so, if they have experienced any difficulty obtaining the data from these authors.

Response: As can be seen in our response to point number 13, more information on the timeline of our project has now been provided in the *'project and data management'* section. Regarding the difficulties experienced thus far in obtaining data, we have now mentioned (page 13):

'As of July 20th 2018, researchers of a total of nine studies have already agreed to be part of the consortium, with two of them having transferred their data already. Of the other studies, application procedures for obtaining the data have started or studies were excluded from our consortium. Reasons for exclusion were the inability to get in contact with researchers (in some cases researchers have retired or have passed away) or the researchers were willing or able to participate.'

Minor points:
Page 9, line 10: "send" should be "sent".
Page 9, line 11: "reminding" should be "reminder".
Page 9, line 48: "study" should be "studies".
Page 10, line 46: delete "in order" and add an "s" to "aim".
Page 10, line 50: change "aim" to "aims".
Page 10, line 53: add the words "or not" after "whether".
Page 11, line 6: delete "in order".
Page 11, line 30: change "analysis" to "analyses".
Page 12, line 20: delete "in order".
Figure 3 – "respons" should be "response"

Response: The manuscript has been revised according to all above suggestions