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)	The influence of work ability and smoking on the prognosis of long-duration activity-limiting neck/back pain – a cohort study of a Swedish working population
AUTHORS	Bohman, Tony; Holm, Lena; Lekander, Mats; Hallqvist, Johan; Skillgate, Eva

VERSION 1 – REVIEW

REVIEWER	Line Thorndal Moll Silkeborg Regional Hospital, Diagnostic Centre
REVIEW RETURNED	26-Aug-2021

GENERAL COMMENTS	Overall: important topic and the focus on modifiable prognostic factors is appreciated. Relevant methodological considerations including e.g. the risk of misclassification. However, I suggest that the risk of residual confounding is covered better in the discussion. Also considerations on generalisability of the results should be discussed (almost 40% lost to follow-up. Population of primarily non-manual workers and self-employed) Author affiliations, point 6: suggest the use of capital letters in names page 3 line 10/11 (more numbers than lines!) the confidence interval for RD lacks a . (0.02 i presume) (same in line 39/40 page 11) Abstract and methods: the study must be considered retrospective. Exposure meaurement in 2010 and outcome in 2014. page 8-9: I have not previously seen this operationalisation of work ability. Please explain whether or not you have scientific support in the literature. Also, it would be reasonable to address the validity of the measure work ability (as operationalised here) in the discussion section (since categories seem to be defined for the purpose of this single study). Regarding this, the use of ref. 27 is not considered sufficient since in ref 27, the outcome is sickness absence and not LANBP. And in ref 27, the operationalisation in this manuscript is not used Discussion page 12: you introduce the term 'mental' / 'physical' work ability. I suggest that you use the terminology that matches your operationalisation (poor, moderate and good - then in brackets you can remind the reader how you defined these three categories with respect to physical and mental) Please ensure that the terminology is consequent throughout the manuscript

	Discussion: confidence intervals of the estimates should be given Page 13: I suggest that the choice of merging neck and back pain patients in the same cohort is supported by references describing similarity in prognostic factors Page 14 line 23/24: regarding the stratified analysis. I am aware that self-rated health is used as a proxy of pain intensity. However, it should be described that good selfperceived work ability is beneficial no matter the degree of self-rated health (not no matter the pain intensity) Discussion: the risk of reverse causation is discussed as to pain being a potential confounder. Based on table 1, a large number of potential residual confounders is not addressed although it must be considered an important limitation of the study (only socioeconomic status, migraine and sleep disturbances are adjusted for). There are notable differences in table 1 in e.g. obesity and psycological distress. Discussion, further regarding residual confounding: do you have data on e.g. health literacy? It would be reasonable to hypothesize that non-smokers and people with self-perceived good work ability also are people with good health literacy. Potential residual confounder? Page 14 line 54. An important typo. It should say OUR results (nor or results) Page 15: large confounder control. To my understanding, three confounders are included in the analyses. I would not consider this large confounder control References: appear relevant but rather old. 9 out of 33 are from 2017 or later. The same number (9/33) are from before 2006.

REVIEWER	Esben Flachs Bispebjerg Hospital, Occupational and Environmental Medicine
REVIEW RETURNED	05-Oct-2021

GENERAL COMMENTS	In general a clear and well written manuscript. It is easy to follow and conclusions are fair and balanced. I have a few comments for the authors to consider though. It is stated that the analyses are done on complete cases only, it is however difficult to follow the numbers and how many are excluded because of which reasons. Table 1 on baseline characteristics is made on the entire study population (n=5177), but only around 3300 are included in analyses, predominantly because of missing outcome, but this cannot be definitively seen from the manuscript, as missing on confounders are not stated anywhere, though it might be possible to calculate from table 1. I suggest to include actual information on missing also for confounders. It might also be an idea to only include individuals actually included in the study in table 1 (n around 3300). A supplementary table might then be supplied with the entire population (n= 5177) to provide information on missingness. In a similar vein I cannot determine if crude and adjusted analyses

	are done on the same populations, eg. are both crude and adjusted analyses of DS done on 3,292 individuals, or does the number differ between the two analyses due to missing on confounders. If that is the case it should be stated clearly. And any impact on results from this ought to be considered. The choice of statistical models (general linear regression for main effects and logistic regression for synergistic/interaction) are both valid, but why the change in model between main effects and interaction? Why are exposures reduced to dichotomised versions, for the interaction analysis? With a quite large fraction of individuals with missing on the outcome, I think this warrants some more thoughts. The authors states that it mostly would result in underestimation of results, which is true in the cases stated by the authors. However, more there could be possibilities. What would be the impact of confounders related to the missingness, and could people having (differentially) left the labour market impact he results. Would it be possible to procure information on labour market status or similar, to investigate this in more detail?

REVIEWER	Lingxiao Chen The University of Sydney Institute of Bone and Joint Research
REVIEW RETURNED	24-Oct-2021

GENERAL COMMENTS	1. The authors mentioned a lot about lifestyle factors in the Introduction. However, they only explored smoking in this study. The authors should revise the Introduction to focus on smoking. Some additional references about the potential mechanisms would be preferred. 2. The authors mentioned that they used data from three sub cohorts. More details are needed to describe these three cohorts. Some previous references from these three should be provided if available. 3. The authors did not provide a justification for the classification of the exposure. I have concerns about potential misclassification of the exposure, which might bias the results. Thus, additional sensitivity analysis to classify the exposure is needed. 4. I also have concerns about the population. Refs 19 and 22 did not focus on these cohorts. The authors should provide a more specific ref or perform additional sensitivity analyses to justify the population the authors claimed. 5. The details of confounders should be listed in the main manuscript. 6. Selection confounders based on the statistical estimates are not recommended. The authors should consider a casual diagram. Please check the updated version of Modern Epidemiology (2021). 7. The authors should provide missing data for each variable listed in Table 1. If the missing data is not negligible, complete case analyses might bias the results. The authors should consider other methods to handle the missing data issue, such as multiple imputation. 8. The authors should mention more in the Limitation section, such as smoking is defined by a binary question without dose and history information.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Line Thorndal Moll, Silkeborg Regional Hospital Comments to the Author:
Overall: important topic and the focus on modifiable prognostic factors is appreciated. Relevant methodological considerations including e.g. the risk of misclassification. However, I suggest that the risk of residual confounding is covered better in the discussion. Also considerations on generalisability of the results should be discussed (almost 40% lost to follow-up. Population of primarily non-manual workers and self-employed)

Author response:
Dear Dr. Line Thorndal Moll
Thank you for your through revision of our manuscript. We have considered your comments and hope that you will be satisfied with the changes done together with our responses and explanations. Our response (pages and lines) refers to the "Main document - marked copy". The changed tables can be seen in the "Main document". Table 1 has a decreased size in the revised manuscript due to restrictions from the journal (two pages instead of four).

- To better cover the discussion concerning residual confounding we added a paragraph in the discussion (please, also see the response concerning residual confounding (table 1) and "health literacy", below).
    - *Changes made to the manuscript; page 14, line 287-289.*

- The loss to follow-up is a source of potential bias – mainly selection bias. This has already been discussed in the Discussion: “Non-responders had a significantly higher proportion of smokers and individuals with poor self-perceived work ability than did responders. If most of these individuals would experience a favourable prognosis of their long-duration activity-limiting neck/back pain, a scenario we find unlikely, our results may be overestimated” page 15, line 320-323. Also, risks of bias from misclassification of exposures and outcome, as well as the risk of confounding is discussed, all important for the internal validity, and for that reason also important for the external validity. The aim of the study was to investigate the association between of exposures and outcome and not to survey the prevalence. We have no reason to believe that the associations found in this study should differ between study participants and persons not participating in the study why we consider the generalisability of the associations to be valid.
    - *No changes have been made to the manuscript.*

Author affiliations, point 6: suggest the use of capital letters in names

- *Thank you, we have changed to capital letters.*

page 3 line 10/11 (more numbers than lines!) the confidence interval for RD lacks a . (0.02 i presume) (same in line 39/40 page 11)

- *Of course. Thank you, we have changed to 0.02.*

Abstract and methods: the study must be considered retrospective. Exposure meaurement in 2010 and outcome in 2014.

- We have considered your comment. In general, a cohort study commonly is considered prospective as the population is followed longitudinally and the exposure is measured before the outcome as is the case in the present study. However, we agree that defining a cohort study as retrospective or prospective is not always that easy. There are situations when the term retrospective may be more appropriate to use. This is something discussed at pages 116-117 in Modern epidemiology (1). According to the reference you also need to consider the method for collecting the data, especially concerning the exposure variable and outcome used. If the exposure is based on retrospective data and could be affect by the outcome (disease), the study might be considered retrospective. As none of our exposures were collected retrospectively and would not be affected by the outcome, we are convinced that our study should be considered prospective.

- *No changes have been made to the manuscript.*

page 8-9: I have not previously seen this operationalisation of work ability. Please explain whether or not you have scientific support in the literature. Also, it would be reasonable to address the validity of the measure work ability (as operationalised here) in the discussion section (since categories seem to be defined for the purpose of this single study). Regarding this, the use of ref. 27 is not considered sufficient since in ref 27, the outcome is sickness absence and not LANBP. And in ref 27, the operationalisation in this manuscript is not used

- Thank you for this valuable comment. We agree that the statement in the method section regarding reference 27 may be misleading. Our purpose is only to emphasise that the second WAI item has been used as a single predictive measure in previous studies. Consequently, we have changed the text in the paragraph.
- Regarding the operationalisation of the exposure PWA we have no specific scientific support in the literature, but we have used exactly the same operationalisation in two of our recent studies using self-perceived workability as exposure studying the risk for psychological distress and long-duration activity neck/back pain (the latter published in BMJ Open, 2020) (2, 3).
- To meet your request, we have included a paragraph in the discussion addressing the validity of the exposure PWA, as we agree that it is important. We have also included an additional discussion addressing the validity of the exposure Daily smoking.

  - *Changes made to the manuscript: page 8, line 156-161, and page 14-15, line 305-314.*

Discussion page 12: you introduce the term 'mental' / 'physical' work ability. I suggest that you use the terminology that matches your operationalisation (poor, moderate and good - then in brackets you can remind the reader how you defined these three categories with respect to physical and mental) Please ensure that the terminology is consequent throughout the manuscript

- Thank you for the comment. We introduce the term physical and mental work ability already in the method section but considering your suggestion we have made changes in the manuscript which we believe improve and clarify the statement further, thank you.

  - *Changes made to the manuscript: page 12, line 249-255.*

Discussion: confidence intervals of the estimates should be given

- We have used percentages (%) as an interpretation of the actual estimates (RR) to increase the readability in the first paragraph of the discussion. Including CI as percentages would not be appropriate. An alternative would be to include the actual estimates with CI also in the discussion, but as we already have presented them in the result section, we prefer to keep the paragraph as it is. Hope that you agree.

  - *No changes have been made to the manuscript.*

Page 13: I suggest that the choice of merging neck and back pain patients in the same cohort is supported by references describing similarity in prognostic factors

- As already mentioned in the discussion (page 13-14, line 280-284) we have acknowledged our decision to cluster back- and neck pain patients as questionable as prognostic factors may differ and explained why we considered clustering relevant. If, the estimates would have been different for back pain and neck pain in the "a priory" analysis we would have analysed them separated.
- Despite comprehensive research, evidence for prognostic factors in relation to neck and back pain is sparse. In a 2018 scooping review aiming to identify risk factors, prognostic factors, and comorbidities associated with common spinal disorders, the authors only found evidence for a few risk factors but suggested many factors to be associated to both neck and back pain, i.e. potential risk and prognostic factors, indicating possible similarities between prognostic factors for neck and back pain, reference #6 in the manuscript (4).

- In an earlier BMJ Open publication from our group, using the same data, we have studied risk factors (self-perceived work ability and sleep disturbances) for long term activity limiting low back- and neck pain in persons with occasional neck and/or back pain using the same clustering, based on the same decision (3). We have also clustered neck and backpain, defining it as spinal pain, in several previous publications from our research as we consider such clustering valid in some circumstances but not always (5-8).
- Thus, we prefer not to add any references and keep the paragraph as it is.

    - *No changes have been made to the manuscript.*

Page 14 line 23/24: regarding the stratified analysis. I am aware that self-rated health is used as a proxy of pain intensity. However, it should be described that good selfperceived work ability is beneficial no matter the degree of self-rated health (not no matter the pain intensity)

- *We agree and has changed the text accordingly (page 14, line 299-300).*

Discussion: the risk of reverse causation is discussed as to pain being a potential confounder. Based on table 1, a large number of potential residual confounders is not addressed although it must be considered an important limitation of the study (only socioeconomic status, migraine and sleep disturbances are adjusted for). There are notable differences in table 1 in e.g. obesity and psycological distress.

- This is an important issue and we have tried to be as clear as possible to explain how we have addressed confounders. Please see the 'Statistics' section, page 9, line 183-191, and the appendix listing all the 15 confounders considered. The confounder assessment described were done after careful discussion about if they instead possibly could be intermediators or colliders and performed using the change-in-estimate (CE) method recommended by Mickey/Greenland and Rothman et al. (reference 29 and 30 in the manuscript). As an example, in the assessment of all the 15 potential confounders and PWA as exposure, SES, headache/migraine, and sleep disturbances changed the RR by 5 % or more when introduced one by one in the crude analysis and where therefore adjusted for in the final analysis. Thus, as we considered 15 potential confounders, we believe our method for confounder assessment is a strength in our study, not a limitation.
  We hope that the text referred to above explains how confounders were treated and we prefer to keep the text in the method section as it is. However, we have changed some text in the discussion to be more accurate with what we have done.

    - *Changes made to the manuscript; page 14, line 285-289 and in the first "bullet point" at page 3.*

Discussion, further regarding residual confounding: do you have data on e.g. health literacy? It would be reasonable to hypothesize that non-smokers and people with self-perceived good work ability also are people with good health literacy. Potential residual confounder?

- This is an interesting issue. Unfortunately, we have no data on "health literacy" but we believe "health literacy" is associated to socioeconomic status (SES) and we agree on the importance to consider it in a study like ours. As SES was deemed as a confounder in the analysis with PWA as exposure as well as in the analyses with Daily smoking as an exposure the results are already adjusted accordingly. Therefore, we have not included "health literacy" in the discussion of confounding.
- However, we have followed your recommendation to enhance the part concerning the risk of residual confounding by adding a paragraph in the manuscript.

    - *Changes made to the manuscript; page 14, line 285-289.*

Page 14 line 54. An important typo. It should say OUR results (nor or results)

- *Of course, thank you. Or is changed to our.*

Page 15: large confounder control. To my understanding, three confounders are included in the analyses. I would not consider this large confounder control

- Please see our response above, explaining the method used for cofounder assessment. Even if we only adjusted our analyses using 3 and 4 confounders, we considered a large number of confounders using the CE-method. Consequently, we have changed the text in the discussion where we incorrectly used the term "control" and instead use the term "consider". Hope you find this explanation and change adequate.
    - *Changes made to the manuscript; page 14, line 285-289 and in the first "bullet point" at page 3.*

References: appear relevant but rather old. 9 out of 33 are from 2017 or later.
The same number (9/33) are from before 2006.

- We are pleased that you find our references to be relevant. We also believe that it is important to use recently published studies as references when possible. Other issues of importance regarding the references to be included are if they are original sources and/or high-quality sources even if they are rather old. Due to your comment, we have checked our references again. This resulted in the following decisions which we hope that you find satisfactory even though we decide to keep some rather old references (all reference numbers referred to are from the original manuscript):
    - *We regard reference # 7, 8, 9, 19, 20, 21, 23, 29 and 32 to be original sources and/or high-quality sources that we want to keep in the manuscript even if some of them are rather old.*
    - *We have excluded reference #35 because of changing the text at page 15-16, line 330-333.*
    - *Reference #30 is updated to the 4:th edition from 2021 and reference #29 is excluded as it is covered by #30.*
    - *Furthermore, we tried to find more recent references for our references related to the Work ability index (#10, 11, 12, 14, 25 and 26), but even if there are numerous publications related to the WAI, e.g. reliability and validity of translated WAI versions, we found no publications suitable as substitute. However, we have excluded WAI reference #15 and #24 as they are covered by the other reference related to the statements.*
    - *Changes made to reference list and to reference numbers in the text.*

Reviewer: 2
Dr. Esben Flachs, Bispebjerg Hospital
Comments to the Author: In general a clear and well written manuscript. It is easy to follow and conclusions are fair and balanced. I have a few comments for the authors to consider though.

Author response:
Dear Dr. Esben Flachs
Thank you for your through revision of our manuscript. We have considered your comments and hope that you will be satisfied with the changes done together with our responses and explanations. Our response (pages and lines) refers to the "Main document - marked copy". The changed tables can be seen in the "Main document". Table 1 has a decreased size in the revised manuscript due to restrictions from the journal (two pages instead of four).

It is stated that the analyses are done on complete cases only, it is however difficult to follow the numbers and how many are excluded because of which reasons. Table 1 on baseline characteristics is made on the entire study population (n=5177), but only around 3300 are included in analyses, predominantly because of missing outcome, but this cannot be definitively seen from the manuscript, as missing on confounders are not stated anywhere, though it might be possible to calculate from table 1. I suggest to include actual information on missing also for confounders. It might also be an idea to only include individuals actually included in the study in table 1 (n around 3300). A supplementary table might then be supplied with the entire population (n= 5177) to provide information on missingness.

- Thank you for an important comment. Thanks to your comment we have realised that the wording "complete cases" used in our manuscript is wrong. More appropriate would be to use "complete-subject analyses" as suggested by Lash et al. when using regression analyses involving variables for which the records has missing values (1). We have changed the text in the statistic section accordingly and hope that you agree.
- We have defined the study population by including all persons that we intend to follow from baseline (2010) to follow-up (2014). As table 1 should be a description of the baseline characteristics of the study population we strongly believe that we should report descriptive statistics for the 5,177 persons with long-duration activity limiting NP and/or BP and information on the exposures in 2010. We also need to have non-responders in 2014 in the study population to perform the sensitivity analyses and discuss selection bias. This is a traditional way of defining the study population and hope that you find our explanation acceptable. Please, regarding the individuals actually included in the analyses see our answer to your next comment below.

    - *We have followed your recommendation and up-dated Table 1 with numbers of internal missing. We also changed the text in the statistics section, page 10, line 191-192.*

In a similar vein I cannot determine if crude and adjusted analyses are done on the same populations, eg. are both crude and adjusted analyses of DS done on 3,292 individuals, or does the number differ between the two analyses due to missing on confounders. If that is the case it should be stated clearly. And any impact on results from this ought to be considered.

- We apologise for being unclear.
    - *The number of individuals included in the analyses presented in Table 2 and Table 3 are now included.*

- Compared to the crude analysis there are only 8% missing in the adjusted analyses with PWA as exposure and only 6% missing in the adjusted analysis with DS as exposure (Table 2). We do not believe that such a low proportion off missing would have any impact on the result and refrain from discuss it further in the manuscript – hope that you agree.

The choice of statistical models (general linear regression for main effects and logistic regression for synergistic/interaction) are both valid, but why the change in model between main effects and interaction?

- We understand your concern, but to be sure not to introduce biases in our analyses we have used logistic regression for calculating measures of biological interaction as this is the recommended method when using the EpiNET´s epidemiological tool (9).

Why are exposures reduced to dichotomised versions, for the interaction analysis?

- This is simply because a biological interaction analyses concerning synergetic or additive effects are commonly performed with dichotomised exposures as described in Modern epidemiology, chapter 26 (1). Thus, the EpiNET tool used in these analyses only handle dichotomised exposures.
- Furthermore, using three categories in one of the exposures will complicate the interpretation of the interaction and will also challenge the power of the analyses as using three categories of PWA would result in three additional "exposures" categories.

With a quite large fraction of individuals with missing on the outcome, I think this warrants some more thoughts. The authors states that it mostly would result in underestimation of results, which is true in the cases stated by the authors. However, more there could be possibilities. What would be the impact of confounders related to the missingness, and could people having (differentially) left the labour market impact he results. Would it be possible to procure information on labour market status or similar, to investigate this in more detail?

- We agree that there are more possibilities of the impact of the loss to follow-up. We discussed that if most of the individual's lost to follow-up would experience a favourable prognosis of their long-duration activity-limiting neck/back pain, our results may be overestimated, page 15, line 319-323. If the individual's lost to follow-up would experience a worse prognosis our results may be underestimated. “Confounders related to the missingness” is a scenario we have not discussed before, and we are uncertain what it means. For confounding to be present it shall be associated with the exposure and cause the outcome, not with the exposure and loss to follow-up. Likewise, we have problems understanding what “could people having (differentially) left the labour market impact the result” means in terms of causality. We have tried to understand your concerns and believe that what you refer to as potential biases goes down to if exposure status at baseline differs between those followed and those loss to follow-up which we have handled carefully by the discussion of selection bias. Unfortunately, there are no possibilities to produce information on labour market status or similar.

- *No changes have been made to the manuscript.*

Reviewer: 3
Dr. Lingxiao Chen, The University of Sydney Institute of Bone and Joint Research Comments to the Author:

Author response:
Dear Dr. Lingxiao Chen
Thank you for your through revision of our manuscript. We have considered your comments and hope that you will be satisfied with the changes done together with our responses and explanations. Our response (pages and lines) refers to the "Main document - marked copy". The changed tables can be seen in the "Main document". Table 1 has a decreased size in the revised manuscript due to restrictions from the journal (two pages instead of four).

1. The authors mentioned a lot about lifestyle factors in the Introduction. However, they only explored smoking in this study. The authors should revise the Introduction to focus on smoking. Some additional references about the potential mechanisms would be preferred.

    - We address lifestyle factors in the introduction as we believe that increased understanding of their role as prognostic factors for a favourable outcome is important and suggested as important in literature. Furthermore, or intention with elaborating about lifestyle factors as well as other biopsychosocial factors in the introduction is to cover the current knowledge on the topic – which we believe is highly relevant. As being physically activity is already found to be associated to a favourable prognosis, we find it interesting and important to know if the same counts for non-smokers, especially as smoking is highly associated to many other health problems. Furthermore, our research team has also published studies evaluation physical activity, sleep, BMI and “healthy lifestyle” in relation to the prognosis of neck and back pain but so far not focused on smoking (10-14). We believe we already considered smoking in relation to neck and back pain in the introduction, second paragraph at page 5 and last paragraph at page 6.

        - *Therefore, we prefer to keep most of the introduction as it is, and hope that you agree. However, we have made a small change at page 6, line 107-114 to reduce some of the text concerning other lifestyle factors and their association to self-perceived work ability.*

    - Potential mechanisms for smoking to affect neck and back pain are interesting, but still not well understood as we already stated in the discussion. Searching the literature, we found only a few studies with potential mechanisms. We have added one of them that together with the one ref already in the manuscript, to cover more of suggested mechanisms. Hope that you find it relevant.

        - *Changes made to the manuscript; page 13, line 271-276, and a new reference #32.*

2. The authors mentioned that they used data from three sub cohorts. More details are needed to describe these three cohorts. Some previous references from these three should be provided if available.

- The cohort from 2002, 2006, 2007 and 2010 are described in detail in our manuscript reference #21 (15). The description of the survey from 2014 has not yet been published, but we provide information of responders in 2010 and at the follow-up in 2014 in the manuscript.
- In addition, we have added the questionaries from 2010 and 2014 as supplementary files as requested by the editor. It will be the editor's decision if they should be included as supplementary files or not.
- We believe that this altogether will give the reader the possibility to get sufficient information regarding the cohorts if needed and prefer to keep the text as it is.

    - *No changes have been made to the manuscript.*

3. The authors did not provide a justification for the classification of the exposure. I have concerns about potential misclassification of the exposure, which might bias the results. Thus, additional sensitivity analysis to classify the exposure is needed.

    - Regarding the classification of the exposure PWA we have no specific scientific support in the literature, but we have used exactly the same operationalisation in two of our recent studies using self-perceived workability as exposure studying the risk for psychological distress and long-duration activity neck/back pain (the latter published in BMJ Open, 2020) (2, 3).
    - We share your concerns about potential misclassification of the exposures why we have included a paragraph in the discussion addressing the validity of the exposure PWA, as we agree that it is important. We have also included an additional discussion addressing the validity of the exposure Daily smoking.
    - We do not understand what you mean by "additional sensitivity analysis to classify the exposure", but our decisions concerning the classification are based on clinical considerations, literature, and our previous studies performed a priori to the analyses. Importantly, the classification was decided on before we had the results, and we believe performing additional analysis to find the best classification retrospectively would be inappropriate. We hope that you can accept our explanation.

        - *Changes made to the manuscript: page 14-15, line 305-314.*

4. I also have concerns about the population. Refs 19 and 22 did not focus on these cohorts. The authors should provide a more specific ref or perform additional sensitivity analyses to justify the population the authors claimed.

    - Unfortunately, we do not fully understand this comment.
    - Refs 19 and 22 are included as the authors recommend duration, frequency, and the impact on daily activity limitations to be included in measures/definitions of neck and back pain in general. We are fully aware that they do not focus on the cohorts included in the manuscript – just in NP and BP in general. However, due to your comment, we changed the wording "chronic pain" to neck and back pain in the discussion to be more specific concerning our population. Therefore, we also deleted the former reference # 35 from the manuscript.
    - We believe that it is strongly justified to study the population included in our manuscript as long-term activity limiting neck and back pain affect many persons and constitute a large burden of disease. We do not understand why we need to provide a specific ref or perform sensitivity analyses to further justify or population, nor what kind of sensibility analyses that should be.

        - *Changes made to the manuscript: page 15-16, line 331- 333, and in the reference list.*

5. The details of confounders should be listed in the main manuscript.

- As all confounders and their categories are listed in Table 1, we decided to have the list including the information of how the confounders were measured in a supplementary file to reduce the numbers of tables in the manuscript.
- If the editor decides to include the appendix in the manuscript it is fine with us, otherwise we prefer to keep it as it is.

  - *No changes have been made to the manuscript.*

6. Selection confounders based on the statistical estimates are not recommended. The authors should consider a casual diagram. Please check the updated version of Modern Epidemiology (2021).

   - We strongly agree that causal diagrams are important tools when examine causality. Consequently, we have since many years back used them as additional support to our clinical and literature-based considerations when assessing confounders, mediators, colliders, and effect modifiers in this kind of research. We have of course also done the same when performing this study, even though we do not present them in the manuscript.
   - In the new (and old) version of Modern epidemiology causal diagrams are one tool suggested to use when assessing/selecting confounders, among others. However, in some circumstances the authors also suggest statistical analyses i.e. when there are more potential confounder data available than is possible to adjust for in the regression model. Due to the unique data in the present study, we had the opportunity to include 15 potential confounders, the majority with more than two categories, why it would be inappropriate to include them all in our analyses. Therefore, we used the change-in-estimate method (CE) to achieve a more parsimonious regression model. The CE-method is one of several statistical selection suggested in the up-dated version of Modern epidemiology, page 274-275. A method we consider valid and the one we have used in many previous studies form our research team, i.e. (3, 10, 11, 14, 16).
   - Therefore, we rely on our method for confounder selection and refrain from including a causal diagram in the manuscript.

     - *No changes have been made to the manuscript.*

7. The authors should provide missing data for each variable listed in Table 1. If the missing data is not negligible, complete case analyses might bias the results. The authors should consider other methods to handle the missing data issue, such as multiple imputation.

   - We apologies for being unclear regarding missing data.

     - *Internal missing for each variable is included in Table 1 and the number of individuals included in the crude and adjusted analyses are included in Table 2 and Table 3.*

   - Thank you for focusing on an important issue. We have thoroughly considered your recommendation to perform analyses based on imputed data. Our conclusion is that we prefer not to impute data. We sincerely hope that you may agree on our conclusion. Our conclusion is based on the following arguments referring to Lee et al., p. 578-581 (17):
     - The first decision to make is if imputation would produce more accurate results in terms of improving precision or reducing bias. As we have a large study sample, we strongly believe imputing data will not be necessary to improve the precision why the issue of reducing bias is the most important question in our study.
     - Considering reducing bias by imputing data, Lee et al. state that there is lot to gain by imputing missing in confounders (SES, headache/migraine, and sleep disturbances). But, to perform an imputation, we need another variable that strongly correlates to the variable with missing otherwise the imputation estimates would be biased as well. We only have variables that poorly correlates to the

confounder in our data. Therefore, we think that using poorly correlated variables when imputing data may potentially introduce bias rather than reduce it.

- Compared to the crude analysis there are only 8% missing in the adjusted analyses with PWA as exposure and only 6% missing in the adjusted analysis with DS as exposure. With such a low proportion off missing the gain of imputing would be limited.

    - *No changes have been made to the manuscript.*

8. The authors should mention more in the Limitation section, such as smoking is defined by a binary question without dose and history information.

    - Thank you for the comment. We agree and have extended the section regarding limitations of the measurement and operationalisation of our exposures, PWA and Daily smoking.

        - *Changes made to the manuscript: page 14-15, line 305-314.*

VERSION 2 – REVIEW

REVIEWER	Line Thorndal Moll Silkeborg Regional Hospital, Diagnostic Centre
REVIEW RETURNED	07-Feb-2022

GENERAL COMMENTS	Prospective vs. retrospective: thanks for your ref which I unfortunately do not have access to. According to Clinical Epidemiology (Fletcher), the study is with no doubt retrospective. IF the study had been done prospectively with the aim of the present publication, I would imagine, that the authors had taken measures along the way to avoid the large number of missings and dropouts. These measures were not taken and it seems obvious that the study is done in a retrospective manner owing to the availability of data which the authors find interesting. Line 60-62: I strongly suggest NOT making inferences about causality due to the observational nature of the study The term reverse causation: I suggest using the term 'confounding by indication' if this is what you mean. Line 237-38: no information on neck/back pain at baseline: hence, there is no way to ensure that confounding by indication is not present. How do we know, that participants are not people without neck/back pain and therefore have a more favorable prognosis regarding the outcome in 2014? Line 335-337 I do not follow the argumentation here. Regarding confounding by indication see above. The additional analyses using self rated health do not eliminate this problem.I think the authors' discussion of this issue is unclear. Though generalizability is touched upon, I still miss further considerations. The population is primarily non-manual workers or self-employed. It is mentioned in results section but not in the discussion.

REVIEWER	Esben Flachs Bispebjerg Hospital, Occupational and Environmental Medicine
REVIEW RETURNED	17-Jan-2022

GENERAL COMMENTS	Thank you for the revised manuscript. With the changes proposed I think it is acceptable.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2
Dr. Esben Flachs, Bispebjerg Hospital
Comments to the Author:
Thank you for the revised manuscript. With the changes proposed I think it is acceptable.

Reviewer: 1
Dr. Line Thorndal Moll, Silkeborg Regional Hospital Comments to the Author:

Dear Dr. Line Thorndal Moll
Thank you for your revision of our manuscript. We have considered your comments and hope that you will be satisfied with the changes done together with our responses and explanations. Our response (pages) refers to the "Main document - marked copy".

Reviewer:
Prospective vs. retrospective: thanks for your ref which I unfortunately do not have access to. According to Clinical Epidemiology (Fletcher), the study is with no doubt retrospective. IF the study had been done prospectively with the aim of the present publication, I would imagine, that the authors had taken measures along the way to avoid the large number of missings and dropouts. These measures were not taken and it seems obvious that the study is done in a retrospective manner owing to the availability of data which the authors find interesting.

Author response:
Dear Dr Thorndal Moll, we did address this comment in the former letter referring to Modern epidemiology (Lash et al.). We have attached the chapter referred to as we still believe it supports our description of the design as prospective.
Please also see reference # 20 in our manuscript describing the Stockholm Public Health Cohort (SPHC) from the first survey in 2002 to the 2010 survey (1). In this reference, the purpose of the SPHC was stated as follow: "The cohort is an international resource for epidemiological research, and the data available to the research community for specific studies obtained approval from the Stockholm Public Health Cohort Steering Committee and the Stockholm Regional Ethical Review Board. Furthermore, please also see the description of the SPHC, at the website "Stockholm Public Health Cohort for researchers", including the study design:

"The Stockholm Public Health Cohort is a prospective study set within the framework of the Stockholm County Council public health surveys of 2002, 2006, 2010 and 2014".

Thus, the cohort was followed "along the way", as you mention, from 2002 to 2014. In our study we have used data for the exposure from the 2010 survey and for the outcome from the 2014 survey which ensures that the exposure is independent from the outcome, and measured prior to the "time at risk", and also that the outcome was measured after "time at risk" – all things of importance in a prospective cohort study. The relatively large number of dropouts was probably mainly due to the long period between baseline to follow-up.

Furthermore, we do not understand you referring to Fletcher when classify the study as retrospective. He describes a retrospective (historical) cohort as follows: "Retrospective cohorts are made by going back into the past and assembling the cohort, for example, from medical records, then following the group forward to the present" (Figur 6.2, p. 95). He continues: "Historical cohort studies can take advantage of computerized medical databases and population registries that are used primarily for patient care or to track population health. The major advantages of historical cohort studies over classical prospective cohort studies are that they take less time, are less expensive, and are much easier to do. However, they cannot undertake studies of factors not recorded in computerized databases, so patients' lifestyle, social standing, education, and other important health determinants usually cannot be included in the studies". Thus, in a retrospective cohort you measure the outcome *at present time and the assemble the cohort and collects exposure data retrospectively from medical records and other registries, which is not what we have done, as explained above. As a conclusion we are still convinced that our study has a prospective study design.*

- *No changes made to the manuscript.*

Reviewer:
Line 60-62: I strongly suggest NOT making inferences about causality due to the observational nature of the study

Author response:

Thank you for your suggestion. We strongly believe that the aim of epidemiological studies of this kind and design, i.e. longitudinal observation study, is to contribute to the knowledge of causal inference and that findings supporting inference about causality should be clearly addressed in the text. However, even if clearly addressed you need to be careful not to overstate the findings and discuss possible biases, criteria's we think we have fulfilled. Please see the publication by Professor Miguel A. Hernán which supports or decision to make inferences about causality (2). As a summary, we will keep the C-word, but to meet your suggestion, we have changed the text to be even more careful, not to overstate our findings.

We agree though that the wording in Strength and limitations of the study may be too strong and have changed it accordingly

- *Changes made to the manuscript: page 3, line 58.*

Reviewer:
The term reverse causation: I suggest using the term 'confounding by indication' if this is what you mean.

Author response:
Thank you or your comment. We still think the term reversed causality is more correct than “confounding by indication”. Even the two terms describes the same “type” of bias, the term “confounding by indication” shall be used in observational studies or nonrandomized studies that involves an intervention. For example, in phramaco-epidemiology studies where the effect on the outcome is compared between individuals who take a specific drug and individuals who have not taken the drug. The problem arises from the fact that those who take the drug generally differs from those who do not according to the medical indication for which the drug was prescribed, in other words, individuals are selected for the intervention based on clinical indications. As this is not the case in the present study, i.e. the exposure/independent factor has nothing to do with an intervention, we think the bias we discuss is reversed causality, and should not be expressed as confounding by indication.

- *No changes made to the manuscript.*

Reviewer:
Line 237-38: no information on neck/back pain at baseline: hence, there is no way to ensure that confounding by indication is not present. How do we know, that participants are not people without neck/back pain and therefore have a more favorable prognosis regarding the outcome in 2014?

Author response:

Unfortunately, we do not understand this comment. Firstly, we have no text in the manuscript like the one you mention "no information on neck/back pain at baseline". At page 10 – first paragraph under "Additional analyses" we say that we "had no information on <u>the intensity of LANBP at baseline</u>". All participants included in the cohort reported that they have had low back pain and/or neck pain on average a few days per week or more during the preceding 6 months at baseline (please see the last *paragraph in the "Design and study population" section, page 7-8). Thus, the risk that we have participants without LANBP is limited. Secondly, we do not ensure the absence of reversed causation (the term we prefer), that is why we discuss the issue and argue that the risk is limited and that our additional analysis stratified by good and poor self-rated health supports this statement.*

- *Changes made to the manuscript: We have added text at p. 10, line 200, to further clarify that all participants reported LANBP.*

Reviewer:
Line 335-337 I do not follow the argumentation here. Regarding confounding by indication see above. The additional analyses using self-rated health do not eliminate this problem. I think the authors' discussion of this issue is unclear.

Author response:
Please, se the answer above – we do not think the additional analyses eliminate the problem; it could still be something that affect our results – leading to overestimation – which we clearly have discussed already, but it supports our discussion of a limited risk. However, we have added some text about the issue in the discussion to clarify, hope you find it satisfactory.

- *Changes made to the manuscript: page 20, line 305-307.*

Reviewer:
Though generalizability is touched upon, I still miss further considerations. The population is primarily non-manual workers or self-employed. It is mentioned in results section but not in the discussion.

Author response:
We agree that we cannot be sure about that our results are valid also in a general population with another socioeconomic profile and have added that as a limitation in the Discussion.

- *Changes made to the manuscript: page 21, line 330-332.*