

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) | Are trajectories of neck–shoulder pain associated with sick leave and work ability in workers? A one-year prospective study |
| AUTHORS | Hallman, David; Holtermann, Andreas; Dencker-Larsen, Sofie; Birk Jørgensen, Marie; Nørregaard Rasmussen, Charlotte |

VERSION 1 - REVIEW

| | |
|------------------------|--|
| REVIEWER | Anita Feleus University of Applied Sciences Rotterdam The Netherlands |
| REVIEW RETURNED | 02-Mar-2018 |

| | |
|-------------------------|---|
| GENERAL COMMENTS | <p>The authors performed additional/secondary analyses on cohort data after previous LCGA on NSP-scores of workers over 12 months time. The original cohortstudy focussed on physical activity at work and risks for musculoskeletal pain.</p> <p>In the presented study the predicted value of NSP-trajectory on sick leave and workability is evaluated. Strong is the 4-week monitoring on pain during 12 months to give insight in trajectories of neck-shoulder pain (NSP).</p> <p>Introduction</p> <ul style="list-style-type: none">- It would be informative for the readers to report more specific on incidence/prevalence of NSP and related sick leave/workability and estimated costs of NSP.- Why is it important to examine the predictive validity of pain-trajectories? Or is it important to identify workers at high risk of sick leave? Or at high risk of an unfortunate trajectory regarding pain itself? <p>Methods</p> <p>For details regarding Methods the authors refer to previous publications</p> <ul style="list-style-type: none">-Please make clear whether: 1 the measured 'sick leave' is related to all musculoskeletal complaints or only due to neck-shoulder complaints 2.and whether sick leave is measured in work days or weekdays.-Please be more specific regarding the frameworks or models included possible indicators were based upon . Which models were used?- Please be more specific regarding the measurement of the variable physical load. Which scale? Reference?-Furthermore, one can imagine that autonomy and many other variables may also be important when predicting sick leave and workability. If not included, please discuss it later on in the manuscript. <p>Statistics</p> |
|-------------------------|---|

-Can you please explain what is new in the presented study, since ref. 20 seems to describe the LCGA regarding NSP?
-Was the outcome 'workability' used in GEE dichotomous or continuous? And when used dichotomously, how was the cut-off point determined?

Results

-The number of classes may be a bit confusing for the reader..
In Methods, 4 classes are already mentioned to be included in GEE with the names of the classes Is the LCGA part of the presented study or part of a previous study?
-In results a previous step is described.. the LCGA with originally 6 classes..and the reduction into 4.
In the next paragraph is mentioned that the classes differ in sick leave days and workability. That seems logic taking the clustering of class 1-3 into account.
-Apparently physical load had a separate effect on both outcomes. There are many other possible variables that may have a positive effect on sick leave or workability ..
Please discuss this later in the manuscript.

Discussion

-You describe:

Thus, severe persistent NSP appears to be strongly associated with sick leave due to musculoskeletal pain, regardless of multisite pain and other personal and occupational factors.

In the introduction and discussion no models and available studies are described on (work)variables that do have an effect on sick leave. Please check for relevant studies and add this information.

-In addition to what you report, that NSP trajectories are predictive of sick leave and workability).., it would be clinically interesting to know:

1. which factors can decrease the probability for sick leave and increase workability

2 which variables are associated with the trajectories ..

Because so far, these results do not provide tools for the clinic. Can you report or reflect on that?

General:

In this study secondary analyses are conducted on a cohort of workers that does not seem to have the primary focus on indicators of sick leave or workability.

The info on trajectories of pain is interesting, although the results on LCGA seem to have been reported earlier.(ref 20?).

In the discussion , in my opinion, the manuscript lacks putting the results in broader perspective. The manuscript would improve when providing:

1. insight in the used models and theories

2. a more critical view on variables that are not included and would have been relevant

(ad 8.: therefore i scored no)

3. clear insight in measuring physical load

4. insight in the scale of workability was used as

outcome measure (continuous or high/low, and cut-off point)

| | |
|------------------------|---|
| REVIEWER | Lene Aasdahl Norwegian University of Science and Technology, NTNU Norway |
| REVIEW RETURNED | 07-Mar-2018 |

| | |
|-------------------------|---|
| GENERAL COMMENTS | <p>This manuscript assesses the associations between latent trajectories of neck-shoulder pain (NSP) and self-reported sick leave and workability. As sickness absence due to musculoskeletal complaints is a vast problem, trying to reveal sub-populations who need different treatment approaches is of great interest, and hence the topic is interesting. However, I do have some comments I think should be addressed before this manuscript can be considered for publication.</p> <p>Major issues:</p> <ul style="list-style-type: none"> - Pain and sick leave were measured at the same time-points. It is therefore not surprising that the authors find an association between the two variables. Participants may also, consciously or unconsciously, report more pain if they have sickness absence that month to “justify” their absence. The authors have not discussed the limitations with performing analyses this way in the manuscript. -The trajectories are based solely on pain. In the Discussion, the authors write that the association with sick leave and work ability support the predictive validity of the trajectories, but as the measures are taken at the same time this argument does not hold very well. In a previously published article, the authors found that the trajectories differed in baseline characteristics. Could it not be that pain is just one factor describing these subgroups of NSP patients? We know from numerous previous studies that pain, function and sick leave is complex. I recommend including this in the discussion. -As the results stand today, I do not see them add much to this field. Have the authors considered lagged analyses? This might make the results more interesting. I also wonder whether the authors have investigated the association between the baseline pain scores and sick leave? This association might be equally strong. -It is not described how the work ability variable was included in the analyses. It seems to be dichotomized, but this should be more clearly stated. It should also be described how many participants that were in each category. - The heading “Effect of NSP trajectory on sick leave” and similar for workability is in my opinion misleading as this study do not evaluate the effect but rather the association between the NSP trajectories and the sick leave/workability. For the same reason the authors should reconsider the wording in the title (“influence on sick leave...”) and also in the abstract («...sick leave and workability were affected by...»). -Relative risk for number of days on sick leave each month is not easy to interpret for the reader and needs a more thorough explanation. Adding an absolute estimate, like for examples number of sick leave days would make the results easier to interpret in terms of clinical significance. <p>Minor issues:</p> <ul style="list-style-type: none"> -In the first paragraph of the introduction, neck-shoulder pain is described as «a disorder», while in the second paragraph a «heterogeneous condition». The authors should be more |
|-------------------------|---|

| | |
|--|--|
| | <p>consistent in their choice of words; the latter is perhaps a better description in this case.</p> <p>-It is not clear how many and what type of workplaces that were included in the study, and how the workers at these workplaces were recruited. This is necessary information for the reader to be able to assess the generalizability of the results.</p> <p>- Occupational sector is described as blue collar or administrative/office work in the confounder section on page 7, but under Statistical analyses it is stated that the analyses were performed adjusted for occupational sector made up of four categories.</p> <p>-The median number of sickness absence days for each NSP trajectory would be of interest and the authors should include this in the results section (in addition or instead of supplement figure B).</p> <p>Table 1:</p> <p>-As the number of sick days is not normally distributed, a median (with 25 and 75th percentiles) should be presented in addition to the mean.</p> <p>- The abbreviation NSP should be explained.</p> <p>Table 2:</p> <p>- Information about number of participants in each category of the NSP trajectories are missing.</p> <p>-It is not clear what the OR is describing, the OR for a workability score under 7? This also applies to table 3.</p> <p>Supplemental figure B and C:</p> <p>These figures are descriptive and should therefore not include standard errors, but rather standard deviations (see the STROBE Statement). As data appear to be asymmetrically distributed, median and percentile range might be more appropriate.</p> |
|--|--|

| | |
|------------------------|--------------------------------------|
| REVIEWER | Alex Burdorf Erasmus MC Rotterdam |
| REVIEW RETURNED | 22-Mar-2018 |

| | |
|-------------------------|--|
| GENERAL COMMENTS | <p>The novelty of this study is the high frequency of measuring presence of neck-shoulder pain, which captures trajectories better than most previous studies on trajectories of MSD. However, a few issues require attention:</p> <p>1. abstract: How are RR linked to days/months sick leave, surely not per 1 day/month? Same is true for work ability, I assume that a dichotomized outcome was used.</p> <p>2. introduction: I was expecting also some considerations as to duration of a spell of pain. Trajectories may help to give better insight into patterns of duration of spells.</p> <p>3. Statistical analysis:</p> <p>- In most trajectory analyses the trajectory class is the outcome variable. Here, it is the independent variable, which needs a bit of a discussion, since trajectories are based on 14 waves, whereas the outcome sick leave is time-varying between waves...Thus, waves 2 - 14 on NPS information are used to 'predict' sick leave in wave 1.</p> |
|-------------------------|--|

| | |
|--|--|
| | <p>- the within-subject analysis is not entirely clear to me...is this a mean-difference fixed effects model? or is this a random model with both within- and between factors?</p> <p>4. Results</p> <p>- the average worker responded to > 90%...this is a very awkward expression...</p> <p>- table 2: I am a bit lost here: (1) OR for work ability suggest a dichotomize variable, and not 01-0 variable in in the table title, (2) does severe NSP really links to 17.6 days on sick leave per months? which in fact would be strange since waves have 4 weeks intervals, and this would imply 17.6 days out of max 20 work days, thus, in essence, these persons are on continuous sick leave throughout the 12 months, and (3) please include number of persons per trajectory.</p> <p>5. Discussion</p> <p>- predictive validity; tricky term given the data collection with trajectory status determined with measurements also collected after the outcome of interest. This also implies to statements as "LCGA to identify sub-populations". The latter is more tricky then presented, since any physician would not wait for 12 months information to act upon persons who apparently are almost on sick leave continuously.</p> <p>- see previous remarks with exact interpretation of the findings with respect to number of sick days, and work ability</p> <p>References:</p> <p>- not all references have complete information, eg ref 15</p> <p>- style of journal presentation differs</p> <p>- given the topics, the following references may provide interesting information: Miedema et al Phys Ther 2016;96:972-84. Feleus et al. Occup Environ Med 2017;74:114-22.</p> |
|--|--|

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Anita Feleus

Institution and Country: University of Applied Sciences Rotterdam
The Netherlands

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

The authors performed additional/secondary analyses on cohort data after previous LCGA on NSP-scores of workers over 12 months time. The original cohortstudy focussed on physical activity at work and risks for musculoskeletal pain.

In the presented study the predicted value of NSP-trajectory on sick leave and workability is evaluated. Strong is the 4-week monitoring on pain during 12 months to give insight in trajctories of neck-shoulder pain (NSP).

Authors: We would like to thank the reviewer for noting the strength of using frequent measurements of pain over 12 months.

Introduction

- It would be informative for the readers to report more specific on incidence/prevalence of NSP and related sick leave/workability and estimated costs of NSP.

Authors: We have revised the manuscript by reporting prevalence of NSP and sick leave, and estimated costs. See revisions on page 4, first paragraph.

- Why is it important to examine the predictive validity of pain-trajectories?

Or is it important to identify workers at high risk of sick leave? Or at high risk of an unfortunate trajectory regarding pain itself?

Authors: We agree that the importance of examining the predictive value of pain trajectories needs a better explanation. Sick leave and reduced work ability are important clinical and occupational outcomes associated with high societal and organization economic costs. Previous studies on latent trajectories of pain have usually compared the identified sub-populations in baseline characteristics and found that the identified subgroups differ in a wide range of factors, including occupational and pain characteristics. However, the potential impact of a specific pain trajectory on the prognosis is largely unknown. Thus, the clinical and occupational relevance of different trajectories identified using data driven procedures such as LCGA are uncertain. It is of general importance for researchers, practitioners and clinicians to understand how different time patterns and levels of pain intensity are associated with the risk of poor outcomes. We also believe that it is important to identify workers at high risk of sick leave due to pain. However this would require a different approach by including other factors apart from pain intensity, which was not within the scope of the present study focusing on sub-groups of NSP.

We have revised the introduction by elaborating on the importance of examining the predictive value of pain trajectories. See revisions on page 4, paragraph 2. Also, we have removed the term “predictive validity” from the entire manuscript, and now refer to “association” or in some cases “predictive value”.

Methods

For details regarding Methods the authors refer to previous publications

-Please make clear whether:

1 the measured 'sick leave' is related to all musculoskeletal complaints or only due to neck-shoulder complaints Authors: The question on sick leave was due to all musculoskeletal complaints. This has been clarified on page 7, paragraph 3.

2. and whether sick leave is measured in work days or weekdays. Authors: we assessed sick leave using text message administered on Sundays, with a reminder the following day. The question about sick leave the past month (page 6) did not differentiate between work days and weekends. This is mentioned as a limitation as it may have resulted in imprecise estimates. See revisions on page 17, paragraph 2.

-Please be more specific regarding the frameworks or models included possible indicators were based upon. Which models were used? Authors: The predictors and outcomes were selected based on theoretical assumptions and empirical studies. This has been incorporated in the introduction (page 4, last paragraph) and methods (page 7, confounders and effect modifiers).

- Please be more specific regarding the measurement of the variable physical load. Which scale? Reference? Authors: We have explained the measurement of physical load, and added a reference on the scale, which was modified from the Borg CR10 scale. This is a commonly used instrument to measure various perceptions, including perceived exertion, and has reliable and valid psychometric properties.

-Furthermore, one can imagine that autonomy and many other variables may also be important when predicting sick leave and workability. If not included, please discuss it later on in the manuscript. Authors: We agree that autonomy and many other variables may be important to predict sick leave and work ability. However, this study focused on the pain trajectories, and thus we did not attempt to construct a best prediction model to identify other predictors of sick leave and work ability. We have revised the manuscript by elaborating on the multifactorial nature of sick leave and work ability in the introduction (page 4, last paragraph) and discussion (page 17, last paragraph).

Statistics

-Can you please explain what is new in the presented study, since ref. 20 seems to describe the LCGA regarding NSP?

Authors: The previous study aimed to identify trajectories of neck-shoulder pain (NSP) over 1 year, and to determine whether these trajectories are predicted by NSP characteristics and personal and occupational factors at baseline. The current study uses the same pain trajectories, but has unique aims, research questions, analyses and results, which aim to determine the extent to which the identified trajectories are associated with changes in sick leave and work ability (assessed frequently during one year). Thus, while the previous study focused on sub-groups of NSP and their potential predictors at baseline, this study is novel in investigating how the pain trajectories associates with the risk for sick leave and work ability over time.

-Was the outcome 'workability' used in GEE dichotomous or continuous? And when used dichotomously, how was the cut-off point determined?

Authors: The outcome work ability was used as an ordinal variable, and thus we specified a multinomial distribution for this outcome. This is now clarified in the statistical analyses section, page 9 paragraph 2. The cut point denoting poor work ability was only used for interpretation of the results (e.g. Fig 2b), since it was not feasible to illustrate the association with pain intensity with all 11 categories of work ability. The cut-point is based on Neupane et al. 2011 and Gould 2008 (page 7, paragraph 4).

Results

-The number of classes may be a bit confusing for the reader..

In Methods, 4 classes are already mentioned to be included in GEE with the names of the classes Is the LCGA part of the presented study or part of a previous study?

-In results a previous step is described.. the LCGA with originally 6 classes..and the reduction into 4.

In the next paragraph is mentioned that the classes differ in sick leave days and workability. That seems logic taking the clustering of class 1-3 into account.

Authors: We agree that the number of classes may be confusing. The six classes were identified using LCGA and presented in a previous publication as described above. Merging of the three classes with low pain, high work ability and low sick leave was made to obtain a larger reference group. It did not make sense to determine associations with sick leave and work ability separately in these classes with very low NSP. We have revised the manuscript by removing parts of the methods

describing the number of classes to the results. See revisions on page 9 last paragraph, and page 14, first paragraph).

-Apparently physical load had a separate effect on both outcomes. There are many other possible variables that may have a positive effect on sick leave or workability ..

Please discuss this later in the manuscript.

Authors: We have now elaborated in the discussion on other factors potentially affecting sick leave and work ability. See revisions on page 17, last paragraph.

Discussion

-You describe:

Thus, severe persistent NSP appears to be strongly associated with sick leave due to musculoskeletal pain, regardless of multisite pain and other personal and occupational factors.

In the introduction and discussion no models and available studies are described on (work)variables that do have an effect on sick leave. Please check for relevant studies and add this information.

Authors: Thank for pointing this out. We have now added text and references on work factors affecting sick leave in the introduction (page 4) and discussion (page 17, last paragraph). We also discuss our primary findings in relation to previous studies (page 15, last paragraph, and page 16 second paragraph).

-In addition to what you report, that NSP trajectories are predictive of sick leave and workability) .., it would be clinically interesting to know:

1. which factors can decrease the probability for sick leave and increase workability
- 2 which variables are associated with the trajectories ..

Because so far, these results do not provide tools for the clinic.

Can you report or reflect on that?

Authors: 1) we agree that it is relevant to know more about factors that can decrease the probability of sick leave and increase work ability. However, we believe that this would require extensive analyses and reporting of additional models, which would very much widen the scope of the study and make the manuscript less focused. 2) The variables associated with pain trajectories are presented in a previous study (Hallman et al. SJWEH, 2018). Results from this study are now briefly described in the introduction on page 5, paragraph 3.

Also we believe that the results are relevant for clinicians as they provide novel information about the likelihood of reporting sick leave and reduced workability in sub-groups of workers with different patterns of pain. Importantly our results suggest that in a working population, those with severe persistent pain are much more likely to have sick leave and poor workability than those reporting low NSP, moderate NSP, and strong fluctuating NSP. This result persisted when accounting for baseline pain intensity and multisite pain suggesting that the time pattern of NSP is of clinical and occupational relevance. Thus, workers with signs of more severe persistent symptoms may need special attention regarding prevention and treatment.

General:

In this study secondary analyses are conducted on a cohort of workers that does not seem to have the primary focus on indicators of sick leave or workability.

The info on trajectories of pain is interesting, although the results on LCGA seem to have been reported earlier.(ref 20?). Authors: It is correct that the pain trajectories and their relation to baseline factors are reported earlier (ref 20: Hallman et al 2018, SJWEH), as stated in the introduction.

In the discussion , in my opinion, the manuscripts lacks putting the results in broader perspective. The manuscript would improve when providing:

1. insight in het used models and theories
2. a more critical view on variables that are not included and would have been relevant
(ad 8.: therefore i scored no)
3. clear insight in measuring physical load
4. insight in the scale of workability was used as outcomemeasure (continuous or hogh/ow, and cut-off point)

Authors: Thanks for these suggestions on how to improve the manuscript.

1. We have revised the introduction by providing more information on the relationship between occupational factors (physical and psychosocial), work ability and sick leave to provide a more comprehensive empirical and theoretical background (see also our responses above).
2. The discussion now attempts putting the results in a larger perspective and elaborate on variables (predictors) not included as a limitation of the study.
3. The measurement of physical load is now better described in the methods.
4. The scale of work ability is now clearly explained in the methods and results, with appropriate references to the cut point 7, which was merely used for interpreting the findings.

Reviewer: 2

Reviewer Name: Lene Aasdahl

Institution and Country: Norwegian University of Science and Technology, NTNU
Norway

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

This manuscript assesses the associations between latent trajectories of neck-shoulder pain (NSP) and self-reported sick leave and workability. As sickness absence due to musculoskeletal complaints is a vast problem, trying to reveal sub-populations who need different treatment approaches is of great interest, and hence the topic is interesting. However, I do have some comments I think should be addressed before this manuscript can be considered for publication.

Major issues:

- Pain and sick leave were measured at the same time-points. It is therefore not surprising that the authors find an association between the two variables. Participants may also, consciously or unconsciously, report more pain if they have sickness absence that month to “justify” their absence. The authors have not discussed the limitations with performing analyses this way in the manuscript. Authors: we agree about this limitation and have thus revised the manuscript accordingly by mentioning this issue as a limitation in the methodological discussion (pages 17, paragraph 2).

The trajectories are based solely on pain. In the Discussion, the authors write that the association with sick leave and work ability support the predictive validity of the trajectories, but as the measures are taken at the same time this argument does not hold very well. In a previously published article, the authors found that the trajectories differed in baseline characteristics. Could it not be that pain is just one factor describing these subgroups of NSP patients? We know from numerous previous studies that pain, function and sick leave is complex. I recommend including this in the discussion.

Authors: we agree that other factors differ between these sub groups, and may thus contribute to sick leave and work ability. As shown previously (Hallman et al 2018, SJWEH), the pain characteristics at baseline explained a substantial proportion of variance in pain trajectories, while other factors contributed much less. Physical work load was strongly associated with sick leave among the occupational factors, and it was thus accounted for in this study both as a moderator and a confounder (along with several other potential confounders). Still, there is always a possibility of residual confounding by non-measured factors, which is now included as a limitation of the study. See revisions on page 17, last paragraph. We have included text in the discussion elaborating on multiple factors that may contribute to sick leave and workability (page 17, last paragraph). We have also extended the introduction to emphasize the multifactorial causes of sick leave and work ability (page 4).

-As the results stand today, I do not see them add much to this field. Have the authors considered lagged analyses? This might make the results more interesting. I also wonder whether the authors have investigated the association between the baseline pain scores and sick leave? This association might be equally strong. Authors: We believe that the results add important knowledge to this field of research. First, there are very few studies on latent pain trajectories (and other sub-group analyses) of neck-shoulder pain, and generally few studies among working populations. Most previous studies have been conducted on patients with low back pain. Second, very few studies, if any, on latent pain trajectories have examined prospective associations with prognostic outcomes such as work ability and sick leave. This information is important for understanding the extent to which the identified subgroups are meaningful and of clinical and occupational relevance. Our findings seem of particular importance since the associations of NSP trajectories with sick leave and workability persisted even when adjusting for baseline pain intensity.

We decided not to perform lagged analyses since the relationship between pain and sick leave the following month(s) is less obvious than the relationship between pain and sick leave the same month. We did not expect any marked lagged effect since it can be expected that workers announce sick leave close in time to an elevation in pain; which was supported by the within person analysis of temporal changes in pain and sick leave. Still, as we assessed the predictor and outcome during the same year, causal inferences should be made with caution. Thus, we have mentioned this as a possible limitation (page 17, paragraph 2).

-It is not described how the work ability variable was included in the analyses. It seems to be dichotomized, but this should be more clearly stated. It should also be described how many participants that were in each category.

Authors: workability was included as an ordinal variable in the regression. It was only dichotomized to illustrate the effect on workability in figure 2b. It is not feasible to plot estimated values from an ordinal

regression having 11 levels of Y. We have explained this more clearly in the revised manuscript: statistical analyses (page 8) and results (page 13).

- The heading "Effect of NSP trajectory on sick leave" and similar for workability is in my opinion misleading as this study do not evaluate the effect but rather the association between the NSP trajectories and the sick leave/workability. For the same reason the authors should reconsider the wording in the title ("influence on sick leave...") and also in the abstract («...sick leave and workability were affected by...»).

Authors: we agree that the wording effect and influence can be misunderstood. The word effect refers to statistical effect, not causal effect. We agree that this study design cannot determine true causality; thus we have removed the word influence from the title.

-Relative risk for number of days on sick leave each month is not easy to interpret for the reader and needs a more thorough explanation. Adding an absolute estimate, like for examples number of sick leave days would make the results easier to interpret in terms of clinical significance. Authors: We have included absolute estimates of sick leave days in figure 1 and in text, and clearly refer to these in the results of relative risk. However, for work ability the estimates of absolute risks are not feasible to present due to the multinomial distribution of the outcome. Thus, we refer to the median work ability in each trajectory class of NSP for absolute numbers. See our revisions in the results on page 12 and page 13.

Minor issues:

-In the first paragraph of the introduction, neck-shoulder pain is described as «a disorder», while in the second paragraph a «heterogeneous condition». The authors should be more consistent in their choice of words; the latter is perhaps a better description in this case.

Authors: We have revised the introduction according to the reviewer (see page 4, first sentence).

-It is not clear how many and what type of workplaces that were included in the study, and how the workers at these workplaces were recruited. This is necessary information for the reader to be able to assess the generalizability of the results.

Authors: The study sample was recruited from 15 workplaces in Denmark (see study design, third sentence). We have now added information about the recruitment (see page 6, study design, first paragraph) – "The initial contact and recruitment of companies were performed in collaboration with a large Danish union. The companies were selected to represent blue-collar occupations with different physical demands at work."

- Occupational sector is described as blue collar or administrative/office work in the confounder section on page 7, but under Statistical analyses it is stated that the analyses were performed adjusted for occupational sector made up of four categories.

Authors: Thanks for pointing out this mistake. We have revised the methods to better describe the variable occupational sector with four categories. We have also included N(%) for cleaning, manufacturing and transportation in table 1.

-The median number of sickness absence days for each NSP trajectory would be of interest and the authors should include this in the results section (in addition or instead of supplement figure B).

Authors: The median (IQR) number of absence days for each NSP category are now included in the revised manuscript (see changes on page 13, paragraph 4). We have thus removed the supplementary figure.

Table 1:

-As the number of sick days is not normally distributed, a median (with 25 and 75th percentiles) should be presented in addition to the mean.

Authors: we now present median and inter quartile range (IQR) in table 1 for skewed data.

- The abbreviation NSP should be explained.

Authors: abbreviations of NSP and BMI are added beneath the table.

Table 2:

- Information about number of participants in each category of the NSP trajectories are missing.

Authors: We have revised the manuscript by including n for each NSP trajectory in table 2.

-It is not clear what the OR is describing, the OR for a workability score under 7? This also applies to table 3. Authors: We have now explained the OR in the text and tables 2-3, which refers to the likelihood of a 1-unit reduction in workability (ordinal scale 0-10).

Supplemental figure B and C:

These figures are descriptive and should therefore not include standard errors, but rather standard deviations (see the STROBE Statement). As data appear to be asymmetrically distributed, median and percentile range might be more appropriate. Authors: we have removed the figures and included median (IQR) in the text.

Reviewer: 3

Reviewer Name: Alex Burdorf

Institution and Country: Erasmus MC Rotterdam

Please state any competing interests or state 'None declared': none declared

Please leave your comments for the authors below

The novelty of this study is the high frequency of measuring presence of neck-shoulder pain, which captures trajectories better than most previous studies on trajectories of MSD. However, a few issues require attention:

Authors: We would like to thank the reviewer for point out the novelty of this study.

1. abstract: How are RR linked to days/months sick leave, surely not per 1 day/month?

Same is true for work ability, I assume that a dichotomized outcome was used.

Authors: The RR indicates the relative increase in the number of days on sick leave per month. We have revised the abstract to make this clearer, and added absolute estimates (i.e. the 1-year estimated mean of sick leave days/month) to the relative risk estimates. The OR for workability indicates the likelihood of a 1-unit reduction in work ability (scale 1-10). We used a multinomial distribution (11 response categories) of the outcome, and thus it is not possible to directly translate the OR to absolute estimates. We have revised the abstract by adding the 1-year median value of workability to facilitate interpretation of the relative estimates. Please see our revisions in the abstract.

2. introduction: I was expecting also some considerations as to duration of a spell of pain. Trajectories may help to give better insight into patterns of duration of spells.

Authors: We agree that the duration of spells is important. However, as we assessed pain intensity during the past month (over 14 waves) without additional information about pain duration, this information could not be used to examine the duration of spells.

3. Statistical analysis:

- In most trajectory analyses the trajectory class is the outcome variable. Here, it is the independent variable, which needs a bit of a discussion, since trajectories are based on 14 waves, whereas the outcome sick leave is time-varying between waves...Thus, waves 2 - 14 on NPS information are used to 'predict' sick leave in wave 1.

Authors: It is correct that information on NSP during 14 waves was used to distinguish trajectory classes, which were then regressed against sick leave over 14 waves during the same year. We have revised the statistical section to clarify that the independent and dependent variables were assessed during the same year (page 8, last paragraph). The lack of data of the outcomes at later time points is discussed as a limitation (page 17, paragraph 2).

- the within-subject analysis is not entirely clear to me...is this a mean-difference fixed effects model? or is this a random model with both within- and between factors?

Authors: This is a generalized estimation equation model with three fixed effects: time (14 waves), pain intensity between subjects (person mean across time points), and pain intensity within subjects (difference between pain intensity at each time point and the person mean pain intensity across time points). We report results of the within person fixed effects. This is now described more clearly in the statistics section (page 9, last paragraph).

4. Results

- the average worker responded to > 90%...this is a very awkward expression...

Authors: this has been reworded to: "on average, the workers had 1.2 missing responses to pain and sick leave (14 waves) and 0.4 missing responses to work ability (4 waves) (Table 1)".

- table 2: I am a bit lost here: (1) OR for work ability suggest a dichotomize variable, and not 01-0 variable in in the table title, (2) does severe NSP really links to 17.6 days on sick leave per months? which in fact would be strange since waves have 4 weeks intervals, and this would imply 17.6 days out of max 20 work days, thus, in essence, these persons are on continuous sick leave throughout the 12 months, and (3) please include number of persons per trajectory.

Authors: Thanks for pointing out these issues:

1) Due to the ordinal scale (0-10) and predominance of high ratings of workability, we used a multinomial distribution for this variable in the GEE. Thus, the OR reflects the likelihood of reporting a 1-unit decrease in workability. We now present the ordinal scale in the table title, and provide an interpretation of the OR in the results (page 13, paragraph 4).

2) Severe NSP links to 17.6 more days on sick leave compared to the reference group, while the actual number of days for severe NSP is about 1.5 per month. We have revised the results by adding the interpretation of relative risks and providing the absolute estimates of sick leave in each trajectory class (page 12, paragraph 3).

3) We have added N for each trajectory class in Table 2.

5. Discussion

- predictive validity; tricky term given the data collection with trajectory status determined with measurements also collected after the outcome of interest. This also implies to statements as "LCGA to identify sub-populations". The latter is more tricky than presented, since any physician would not wait for 12 months information to act upon persons who apparently are almost on sick leave continuously. Authors: we agree that predictive validity is a tricky term and may not be completely accurate in this study since the predictor and outcome were assessed during the same time period. Thus we have removed this term throughout the manuscript. Regarding the second point, continuous sick leave was very rare in this study, with on average 1.5 days/month in those with severe persistent NSP (although with a wide dispersion between individuals).

- see previous remarks with exact interpretation of the findings with respect to number of sick days, and work ability

References:

- not all references have complete information, eg ref 15

- style of journal presentation differs

- given the topics, the following references may provide interesting information: Miedema et al Phys Ther 2016;96:972-84.

Feleus et al. Occup Environ Med 2017;74:114-22.

Authors: we have corrected the reference list and added the paper by Feleus et al in the introduction and discussion.

VERSION 2 – REVIEW

| | |
|------------------------|---|
| REVIEWER | Lene Aasdahl Norwegian University of Science and Technology, NTNU Norway |
| REVIEW RETURNED | 07-Jul-2018 |

| | |
|-------------------------|--|
| GENERAL COMMENTS | <p>The revised version of the manuscript is improved and clearer than the original manuscript. However, I still have some comments I think should be addressed before the manuscript can be considered for publication.</p> <p>1) I still find the use of the word "effect" in the manuscript problematic, e.g. "Effect of NSP trajectory classes on sick leave". The authors write in their response that "The word effect refers to statistical effect, not causal effect"- however, this can easily be misunderstood by readers and interpreted as a causal relationship. As the authors themselves write that the word effect can be misunderstood, I suggest changing it.</p> <p>2) The authors highlight that the estimated number of days on sick leave was nearly 14 times larger for severe persistent NSP compared to low NSP. As the numbers of estimated days are quite small, 0.1 and 1.5/per month, this focus on 14 times larger appears bit strange. The estimates are also quite close for low NSP and moderate NSP (0.0 and 0.3 days respectively- one can wonder if there really is a gradient across the trajectories, or more</p> |
|-------------------------|--|

| | |
|--|---|
| | <p>a question about severe vs not severe pain. (In the abstract, the estimate for low NSP is 0.1, in the Result section 0.0- typing mistake?).</p> <p>3) The estimated days of sick leave lack precision estimates (confidence intervals) in the text and figure 1.</p> <p>4) The authors write in the Discussion: "However, as the causes of sick leave and poor work ability are likely multifactorial, the focus on pain trajectories as predictors is a potential limitation". Which I agree fully with, and why I find it a bit confusing/problematic that the authors in the Conclusion write that the "findings indicate that LCGA can be used to identify important sub-populations of workers with a poor prognosis of sick leave and work ability". I suggest the authors shift the focus to what they wrote in their response to reviewer 1 "the clinical and occupational relevance of different trajectories identified using data driven procedures such as LCGA are uncertain. It is of general importance for researchers, practitioners and clinicians to understand how different time patterns and levels of pain intensity are associated with the risk of poor outcomes". The latter is somewhere this study can contribute.</p> <p>5) Presenting IQR as one number is less informative than presenting the range, i.e. the 25th and 75th percentile.</p> <p>6) Table 1: work ability and NSP are reported like "mean workability" followed by the median and mean values, which can be a bit confusing. Perhaps delete the word "mean" before workability?</p> <p>7) Figure 1: I could not find the figures in the revised manuscript, so I assume they are the same in the original manuscript. I suggest changing the legend for figure 1 (it seems to be the raw output from the software with two listings for all trajectories).</p> |
|--|---|

| | |
|------------------------|---|
| REVIEWER | Peter Herbison University of Otago New Zealand |
| REVIEW RETURNED | 15-Oct-2018 |

| | |
|-------------------------|--|
| GENERAL COMMENTS | <p>I found this paper interesting and have relatively minor comments.</p> <p>I have some problems with the last sentence in the abstract where it says that "LCGA can be useful to distinguish important target groups in future observational/intervention studies on NSP". I am not so sure that having to do 14 phases of data collection over a year would be all that useful. It may be of interest to check that the classification found in this study can be repeated in others, then present characteristics of the groups so that people can be classified without having to wait for a year. I should say that the conclusion in the paper is more reasonable.</p> <p>Is there any indication that there is a healthy worker effect in this population? People with very bad NSP may well have moved on to an occupation where they don't get so much pain. If this is true this study may not be as useful as it could be in identifying people in whom an intervention may be worthwhile.</p> |
|-------------------------|--|

| | |
|--|---|
| | <p>It is not completely clear that the authors previous publication (ref 30) uses the same population with only a different second stage to the analysis. Should this be clearer?</p> <p>Should the identification of the Latent Gold software have a country (USA) as well as one of the states?</p> <p>In the LCGA models was time only entered as a linear term? If it is should this be clear? Does this have any implications for interpretation?</p> <p>The use of one-way ANOVA is not mentioned in the methods.</p> |
|--|---|

| | |
|------------------------|--|
| REVIEWER | Peter Knapp University of York, UK. |
| REVIEW RETURNED | 19-Oct-2018 |

| | |
|-------------------------|---|
| GENERAL COMMENTS | <p>The authors have done a good job of responding to peer review comments on the previous version of the manuscript (which I did not review).</p> <p>The statistical analyses have been conducted appropriately and are reported clearly.</p> |
|-------------------------|---|

| | |
|------------------------|--|
| REVIEWER | Lisa Lix University of Manitoba, Canada |
| REVIEW RETURNED | 30-Oct-2018 |

| | |
|-------------------------|--|
| GENERAL COMMENTS | <p>I have a number of comments focused primarily on improving the statistical methodology and the reporting of the results of this interesting longitudinal paper. In particular, clarity with which the modeling is described requires attention.</p> <p>First, information about the frequency and percent of missing data over time should be presented. The information about missing data currently provided in Table 1 is not so straightforward because it does not characterize the cohort over time on missed measurements. It is likely that the amount of missing data at later time points will be greater than the amount of missing data at earlier time points, but this is not clear from the results presented. Some explanation about attrition over time, if present, should be provided when discussing the results.</p> <p>There is limited value in reporting mean work ability averaged across all time points (Table 1) when the authors are interested in testing hypotheses about change in work ability over time. One option is to include the baseline and endpoint measurements of work ability in Table 1 (mean, median, interquartile range, standard deviation). Another option is to produce a graph that provides descriptive information about these data over time. The same comment applies to number of days on sick leave.</p> <p>The authors do not appear to have taken account of the fact that latent class membership is subject to misclassification bias. The</p> |
|-------------------------|--|

authors seem to have used a deterministic approach to assign individuals to classes, which will lead to erroneous estimates of the association of class membership with the outcome measures. Specifically, once an individual is classified to the latent class with the highest posterior probability (which is the method I assume has been used for classification), that individual's probability of being in the assigned class is treated as having a value of one. As well, once an individual is assigned to the latent class with the highest posterior probability, the representativeness of all individuals in the class is assumed to be the same because all individuals are assumed to have a probability of one of being in the assigned class. Standard errors can be significantly underestimated because the uncertainty about the latent class membership is ignored in the regression modeling. Moreover, analysis of variance is an incorrect choice to test for differences amongst latent groups, as described in the Results section.

Overall, the Methods section would benefit from a comprehensive overview of the modeling approach, its rationale, and the modeling sequence. For example, there is some information about the models that were fit to the data that appears in the Results section but not in the Methods section; the authors mention that they included an interaction between class and time in the models for sick days in the Results section, but they do not mention this in the Methods section. There is also a lack of justification for some key modeling choices such as the selection of a first-order autoregressive structure for the generalized estimating equation (GEE) models and assumption of a linear trend over time in the outcome measures. How did the authors make these choices?

A Poisson distribution is assumed for the number of missed days per month. How was goodness-of-fit for the Poisson distribution established? What a negative binomial regression model considered to account for extra-Poisson variation?

The methodology to examine the effects of fluctuations in pain intensity scores is challenging to follow, in part because the authors now appear to be treating neck and shoulder pain as an observed (i.e., manifest) variable instead of a latent variable. In this component of the analysis, the authors indicate that they wish to test the association between within-individual variations in pain and "fluctuations in sick leave and work ability" (this emphasis on fluctuations is made again the Results section). However, the GEE models that they have adopted are intended to model the average evolution of these outcome measures, not variation. Non-linear mixed-effects models may be a better choice.

Figure 2: It is not clear why the outcome variable values are plotted on the vertical axis and the explanatory variable values are plotted on the horizontal axis. This seems contrary to convention.

In the Methods section, the description of multi-site pain measurement is included in the sub-section about effect modifiers and confounders, but the authors don't clearly articulate its role in the model development in the statistical analysis sub-section. There is a brief mention of assessing the "robustness" of the models by including the number of pain sites in the model, but a lack of clarity about what this phrase means.

| | |
|--|---|
| | <p>I would appreciate a more explicit description of the latent class growth analysis model. While a relevant reference to a published source is given, the description provided in text left me with some questions about the latent variable measurement model. A pictorial diagram of the latent variable measurement model may be a useful addition to the manuscript to ensure readers have sufficient information to understand the modeling process.</p> <p>A potentially important covariate for the analysis is the presence of comorbid chronic conditions. If comorbidity has not been measured for the cohort, please identify this as a potential limitation of the study.</p> |
|--|---|

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 2

Reviewer Name: Lene Aasdahl

Institution and Country: Norwegian University of Science and Technology, NTNU
Norway

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

The revised version of the manuscript is improved and clearer than the original manuscript. However, I still have some comments I think should be addressed before the manuscript can be considered for publication. Authors: We would like to thank the reviewer for stating that the manuscript has been improved and that it is now clearer.

1) I still find the use of the word “effect” in the manuscript problematic, e.g. “Effect of NSP trajectory classes on sick leave”. The authors write in their response that “The word effect refers to statistical effect, not causal effect”- however, this can easily be misunderstood by readers and interpreted as a causal relationship. As the authors themselves write that the word effect can be misunderstood, I suggest changing it. Authors: We have changed the word effect to association throughout the manuscript, with some exceptions where it is embedded in the term used, e.g. “effect modification”, “within subject effect” and “time effect”.

2) The authors highlight that the estimated number of days on sick leave was nearly 14 times larger for severe persistent NSP compared to low NSP. As the numbers of estimated days are quite small, 0.1 and 1.5/per month, this focus on 14 times larger appears bit strange. The estimates are also quite close for low NSP and moderate NSP (0.0 and 0.3 days respectively- one can wonder if there really is a gradient across the trajectories, or more a question about severe vs not severe pain. (In the abstract, the estimate for low NSP is 0.1, in the Result section 0.0- typing mistake?). Authors: We agree that the absolute estimates of sick leave days are rather small, although 1.5 days per month (18 days per year) for severe NSP appears highly relevant compared with 0.1 days per month (1.2 days per year) for low NSP. We have revised the text by removing “14 times larger” in the abstract, results (page 10) and discussion (page 14).

We are not sure how to interpret what the reviewer means with the comment about “gradient across trajectories”? We agree that the estimates are close between low and moderate NSP, although this difference was statistically significant (Table 2). The sick leave estimates for each trajectory class was 1 (low), 3 (moderate), 8 (strong) and 14 (severe), so it appears that there is an additional increase in relative risk for each class (e.g. the difference in estimates between strong and severe NSP is larger than the difference between moderate and strong NSP). This could be due to higher pain intensities, but also due to different patterns of pain in different classes (e.g. severe persistent NSP showed less fluctuation than strong NSP).

Thanks for noting the typing mistake. This is now corrected in the results to 0.1.

3) The estimated days of sick leave lack precision estimates (confidence intervals) in the text and figure 1. Authors: we have added 95% CIs in the text. However, it was not possible to obtain the 95% CIs of the predicted values for each time point. Thus, we hope that it is sufficient to report CIs for the average estimate in the text.

4) The authors write in the Discussion: "However, as the causes of sick leave and poor work ability are likely multifactorial, the focus on pain trajectories as predictors is a potential limitation". Which I agree fully with, and why I find it a bit confusing/problematic that the authors in the Conclusion write that the "findings indicate that LCGA can be used to identify important sub-populations of workers with a poor prognosis of sick leave and work ability". I suggest the authors shift the focus to what they wrote in their response to reviewer 1 "the clinical and occupational relevance of different trajectories identified using data driven procedures such as LCGA are uncertain. It is of general importance for researchers, practitioners and clinicians to understand how different time patterns and levels of pain intensity are associated with the risk of poor outcomes". The latter is somewhere this study can contribute.

Authors: We agree with the reviewer and have therefore revised the conclusion to be less focused on the LCGA approach and more focused on the importance of understanding the consequences of different time patterns and levels of pain, as well as the need for prevention of severe persistent pain. Please see our revisions in the abstract and discussion (page 17):

"This longitudinal study shows that severe persistent NSP is associated with sick leave and reduced work ability over 1 year among workers. The high prevalence of severe persistent NSP and the increase in relative risk of sick leave and poor work ability point to the need for preventive strategies aiming at reducing severe persistent NSP among workers. Overall, our findings contribute with further understanding regarding the possible consequences of different time patterns and levels of NSP, which can be of general importance for researchers, practitioners and clinicians.

5) Presenting IQR as one number is less informative than presenting the range, i.e. the 25th and 75th percentile. Authors: Presenting IQR is commonly used to indicate the dispersion and provides similar information as reporting both the 25th and 75th percentile. We prefer reporting IQR as one number to ease reading and make the table less extensive. However, we will change this if the Editor prefers that format.

6) Table 1: work ability and NSP are reported like "mean workability" followed by the median and mean values, which can be a bit confusing. Perhaps delete the word "mean" before workability?

Authors: We see your point, but also find it important to report that this refers to the mean over all time points. We have revised table 1 in accordance with the suggestion from Reviewer 6 by adding data on work ability and sick leave during baseline and the last follow-up. Thus, we believe that the reporting is clearer now.

7) Figure 1: I could not find the figures in the revised manuscript, so I assume they are the same in the original manuscript. I suggest changing the legend for figure 1 (it seems to be the raw output from the software with two listings for all trajectories).

Authors: we have changed the legend of the figure 1 following the reviewer's comment.

Reviewer: 4

Reviewer Name: Peter Herbison

Institution and Country: University of Otago New Zealand

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

I found this paper interesting and have relatively minor comments.
Authors: Thanks for pointing out that this paper is interesting.

I have some problems with the last sentence in the abstract where it says that "LCGA can be useful to distinguish important target groups in future observational/intervention studies on NSP". I am not so sure that having to do 14 phases of data collection over a year would be all that useful. It may be of interest to check that the classification found in this study can be repeated in others, then present characteristics of the groups so that people can be classified without having to wait for a year. I should say that the conclusion in the paper is more reasonable. Authors: We agree with the reviewer about the conclusion. Thus, we have revised the conclusion in the abstract so that it corresponds to the conclusion in the paper. We have also removed the sentence "...LCGA approach to distinguish important target groups in future..." and emphasize the need for prevention of severe persistent pain.

Is there any indication that there is a healthy worker effect in this population? People with very bad NSP may well have moved on to an occupation where they don't get so much pain. If this is true this study may not be as useful as it could be in identifying people in whom an intervention may be worthwhile. Authors: We agree that a possible healthy worker effect is important to consider. However, due to the high prevalence of severe persistent pain among workers within different occupational sectors, we don't expect a prominent healthy worker effect in this population. This is also supported by the rather high seniority in the job (average 14 years, SD 10 years), which was comparable across NSP trajectories (e.g. 14 years for low NSP and 16 years for severe persistent NSP). Thus, we believe that it is important with future intervention studies aiming at prevention of severe persistent pain in this working population, and have written this more clearly in the discussion section.

It is not completely clear that the authors previous publication (ref 30) uses the same population with only a different second stage to the analysis. Should this be clearer? Authors: we have revised the manuscript by explicitly stating that the present study uses the same population in the introduction and methods. See revision on page 5 and 8.

Should the identification of the Latent Gold software have a country (USA) as well as one of the states? Authors: this has now been corrected.

In the LCGA models was time only entered as a linear term? If it is should this be clear? Does this have any implications for interpretation? Authors: we tested both linear and quadratic models but decided for the linear model since a quadratic term did not improve model fit. This is now added to the methods. See revisions on page 9, first paragraph.

The use of one-way ANOVA is not mentioned in the methods. Authors: we have removed the ANOVA results based on suggestions from Reviewer 6.

Reviewer: 5

Reviewer Name: Peter Knapp

Institution and Country: University of York, UK.

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

The authors have done a good job of responding to peer review comments on the previous version of the manuscript (which I did not review).

The statistical analyses have been conducted appropriately and are reported clearly.

Authors: We would like to thank the reviewer for pointing this out.

Reviewer: 6

Reviewer Name: Lisa Lix

Institution and Country: University of Manitoba, Canada

Please state any competing interests or state 'None declared': I declare I have no competing interests.

Please leave your comments for the authors below

I have a number of comments focused primarily on improving the statistical methodology and the reporting of the results of this interesting longitudinal paper. In particular, clarity with which the modeling is described requires attention. Authors: We would like to thank the reviewer for noting the need to improve the description of the statistical methods. To increase clarity of the statistical methodology and reporting of results, we have added text in the methods and results describing the Latent class growth analysis (LCGA) and regression models in more detail. Also, we now clearly refer to our previous study on construction of the LCGA models (ref 30). See also our response to the specific comments below.

First, information about the frequency and percent of missing data over time should be presented. The information about missing data currently provided in Table 1 is not so straightforward because it does not characterize the cohort over time on missed measurements. It is likely that the amount of missing data at later time points will be greater than the amount of missing data at earlier time points, but this is not clear from the results presented. Some explanation about attrition over time, if present, should be provided when discussing the results. Authors: Thanks for pointing out this important aspect of missing data across time. We have added text in the results about the response rate at early and late time points. The amount of missing data increased over time, but to a rather small extent. During the second wave, the response rates were 96%, 96% and 93% for pain, sick leave and work ability, respectively, while the response rates during the last wave dropped to 87%, 88% and 90%. See revisions in the results text on page 10 and Table 1, and discussion on page 16. Specifically we state that missing data over time may have introduced some uncertainties in the models.

There is limited value in reporting mean work ability averaged across all time points (Table 1) when the authors are interested in testing hypotheses about change in work ability over time. One option is to include the baseline and endpoint measurements of work ability in Table 1 (mean, median, interquartile range, standard deviation). Another option is to produce a graph that provides descriptive information about these data over time. The same comment applies to number of days on sick leave. Authors: We agree with the reviewer. Thus we have added values (mean, median, SD, IQR) during baseline and the last follow-up in table 1 for pain, sick leave and work ability.

The authors do not appear to have taken account of the fact that latent class membership is subject to misclassification bias. The authors seem to have used a deterministic approach to assign individuals to classes, which will lead to erroneous estimates of the association of class membership with the outcome measures. Specifically, once an individual is classified to the latent class with the highest posterior probability (which is the method I assume has been used for classification), that individual's probability of being in the assigned class is treated as having a value of one. As well, once an individual is assigned to the latent class with the highest posterior probability, the representativeness of all individuals in the class is assumed to be the same because all individuals are assumed to have a probability of one of being in the assigned class. Standard errors can be significantly underestimated because the uncertainty about the latent class membership is ignored in the regression modeling. Moreover, analysis of variance is an incorrect choice to test for differences amongst latent groups, as described in the Results section. Authors: we agree with the reviewer that misclassification of class membership is important to consider. As presented previously (ref 30) the entropy for the current LCGA was 0.92 and the classification error was fairly small (i.e. 0.06). Thus, it seems unlikely that adjustment for classification errors would markedly change the present results. We have removed the ANOVA from the results.

Overall, the Methods section would benefit from a comprehensive overview of the modeling approach, its rationale, and the modeling sequence. For example, there is some information about the models

that were fit to the data that appears in the Results section but not in the Methods section; the authors mention that they included an interaction between class and time in the models for sick days in the Results section, but they do not mention this in the Methods section. There is also a lack of justification for some key modeling choices such as the selection of a first-order autoregressive structure for the generalized estimating equation (GEE) models and assumption of a linear trend over time in the outcome measures. How did the authors make these choices?

Authors: Thanks for informing us about the inconsistent description of the GEE modeling approach. We have revised the manuscript to increase clarity. The interaction between class and time, and the rationale for excluding this interaction from the model is now described in the methods. Also, we have provided arguments justifying the covariance structure and the assumptions of a linear trend in the outcome measures. See revisions in methods on page 8-9.

A Poisson distribution is assumed for the number of missed days per month. How was goodness-of-fit for the Poisson distribution established? What a negative binomial regression model considered to account for extra-Poisson variation? Authors: We assumed that days on sick leave (counts) followed a Poisson distribution, but did not explicitly test the goodness of fit of expected versus observed data. As suggested by the reviewer, we have remodeled the data to account for extra variation by using a negative binomial regression model. This approach gave similar estimates and confidence intervals as the Poisson model (see Table below for comparisons based on the unadjusted model 1). Thus, we expect our results based on the Poisson distribution to be reasonable.

| | Original model 1 | | | Negative binomial | | |
|--------------|------------------|--------|------|-------------------|--------|------|
| | Exp B | 95% CI | | Exp B | 95% CI | |
| Moderate NSP | 3.3 | 1.9 | 5.7 | 3.3 | 1.9 | 5.6 |
| Strong NSP | 9.0 | 4.8 | 16.9 | 9.0 | 4.8 | 16.9 |
| Severe NSP | 17.6 | 9.4 | 33.2 | 17.5 | 9.3 | 32.8 |

The methodology to examine the effects of fluctuations in pain intensity scores is challenging to follow, in part because the authors now appear to be treating neck and shoulder pain as an observed (i.e., manifest) variable instead of a latent variable. In this component of the analysis, the authors indicate that they wish to test the association between within-individual variations in pain and “fluctuations in sick leave and work ability” (this emphasis on fluctuations is made again the Results section). However, the GEE models that they have adopted are intended to model the average evolution of these outcome measures, not variation. Non-linear mixed-effects models may be a better choice.

Authors: Thanks for pointing out the need for a clearer description of the models of within person effects of pain intensity on sick leave and work ability.

In brief, model 1 included three fixed effects:

- (A) Time (14 waves).
- (B) Person mean pain intensity (i.e. the average value for each person across 14 waves).
- (C) Difference from the person mean pain intensity (i.e. mean centering of the pain intensity scores across waves for each person).

Since the time varying predictor (pain intensity) consists of both between and within subject variability, we adjusted the models for the person mean pain intensity (B) to partition the within subject variability (C). Thus, when including the effects A, B and C in the model, the main effect of C reflects the average within person effect of monthly change in pain intensity on the outcome (sick leave or work ability), with A and B being held constant. We believe that it is correct to interpret this as temporal fluctuations (i.e. variability around the mean across time points) in pain intensity. Still, we agree with the reviewer that temporal fluctuations in the outcomes may be misleading. Thus, we have removed this from the revised manuscript.

Figure 2: It is not clear why the outcome variable values are plotted on the vertical axis and the explanatory variable values are plotted on the horizontal axis. This seems contrary to convention. Authors: we used sick leave and work ability as outcome variables and present those on the y-axis in the figures. This is clarified in the methods of the revised manuscript.

In the Methods section, the description of multi-site pain measurement is included in the sub-section about effect modifiers and confounders, but the authors don't clearly articulate its role in the model development in the statistical analysis sub-section. There is a brief mention of assessing the "robustness" of the models by including the number of pain sites in the model, but a lack of clarity about what this phrase means. Authors: We included multi-site pain in the model since more severe neck-shoulder pain may be accompanied by comorbidity of pain in other sites. Multi-site pain could influence sick leave and work ability. Thus, we wanted to test whether the results from the primary models were consistent when adjusting for multisite pain. We have now revised this sentence in the methods on page 9.

I would appreciate a more explicit description of the latent class growth analysis model. While a relevant reference to a published source is given, the description provided in text left me with some questions about the latent variable measurement model. A pictorial diagram of the latent variable measurement model may be a useful addition to the manuscript to ensure readers have sufficient information to understand the modeling process. Authors: We have described the LCGA modeling and the resulting trajectory classes in more detail in a previous study on the same study sample. We now clearly refer to that study, and, thus, prefer not replicating a more extensive description of the LCGA models. We have slightly revised the description of the LCGA modeling and refer to Jung and Wickrama (2008) and van de Schoot et al (2016) for pictorial diagrams.

A potentially important covariate for the analysis is the presence of comorbid chronic conditions. If comorbidity has not been measured for the cohort, please identify this as a potential limitation of the study. Authors: We agree about this, and have thus included this as a limitation on page 16.

VERSION 3 – REVIEW

| | |
|------------------------|---|
| REVIEWER | Lene Aasdahl Norwegian University of Science and Technology, NTNU Norway |
| REVIEW RETURNED | 09-Dec-2018 |

| | |
|-------------------------|--|
| GENERAL COMMENTS | The authors have done a good job revising the manuscript. The revised version is more focused and clearer. My previous comments are all answered satisfactory. I do not have any further comments. |
|-------------------------|--|

| | |
|------------------------|---|
| REVIEWER | Peter Herbison University of Otago New Zealand |
| REVIEW RETURNED | 02-Dec-2018 |

| | |
|-------------------------|---|
| GENERAL COMMENTS | I have no further comments on this article. |
|-------------------------|---|

| | |
|------------------------|--|
| REVIEWER | Lisa Lix University of Manitoba, CANADA |
| REVIEW RETURNED | 02-Jan-2019 |

| | |
|-------------------------|---|
| GENERAL COMMENTS | <p>The authors have done a good job of responding to the reviewer comments. I have just a few minor revisions for the authors to attend to.</p> <p>(1) In the Statistical Analysis section, the authors use a variety of terminology to describe the regression models that were fit to the data, calling them "GEE", "GEE regression modeling" and "GEE models". I would appreciate some standardization of the language that is used throughout this section and also in the Results section.</p> <p>(b) In Table 1, the abbreviation SMS should be defined in the table note.</p> <p>(c) In Tables 2 and 3, it would be helpful to have quantitative information about the fit of the models that are applied to the sick leave and work ability measures. The authors fit a series of models that are unadjusted for covariates (Model 1), then adjusted for age, sex, and body mass index (Model 2), then adjusted for age, sex, body mass index, occupational sector, and physical work load (Model 3). A quasi-likelihood-based model-selection criterion (QIC) statistic is available to describe the fit of models that include generalized estimating equations (GEEs).</p> <p>(d) In Table 3, the table title indicates that the analysis is for neck-shoulder pain. However, within the table, directly above Model 1, the phrase "NSP intensity" has been included. Could this phrase be removed and the table title modified (as needed) to clarify the associations that are being tested?</p> |
|-------------------------|---|

VERSION 3 – AUTHOR RESPONSE

Reviewer: 6

Reviewer Name: Lisa Lix

Institution and Country: University of Manitoba, CANADA Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below The authors have done a good job of responding to the reviewer comments. I have just a few minor revisions for the authors to attend to.

(1) In the Statistical Analysis section, the authors use a variety of terminology to describe the regression models that were fit to the data, calling them "GEE", "GEE regression modeling" and "GEE models". I would appreciate some standardization of the language that is used throughout this section and also in the Results section.

Authors: We have changed the terminology to make it consistent in the methods and results.

(b) In Table 1, the abbreviation SMS should be defined in the table note.

Authors: We have changed SMS to text messages in Table 1.

(c) In Tables 2 and 3, it would be helpful to have quantitative information about the fit of the models that are applied to the sick leave and work ability measures. The authors fit a series of models that are unadjusted for covariates (Model 1), then adjusted for age, sex, and body mass index (Model 2), then adjusted for age, sex, body mass index, occupational sector, and physical work load (Model 3). A quasi-likelihood-based model-selection criterion (QIC) statistic is available to describe the fit of models that include generalized estimating equations (GEEs).

Authors: We agree that it is relevant to present the fit statistics of the different GEE models. Regarding sick leave, the models for the between-subject effects (Table 2) showed QIC values of 24837 (model 1), 24489 (model 2) and 23686 (model 3) indicating better fit in models with more covariates. Similarly, the models of within-subject effects (Table 3) showed QIC values of 16329 (model 1), 16217 (model 2) and 16015 (model 3).

Regarding work ability, QIC cannot be obtained from the multinomial models because the log quasi-likelihood function can not be derived, and thus, fit statistics are not an option for this GEE model in SPSS. Thus, for consistency across the two outcomes, we prefer not to include QIC statistics in the manuscript.

(d) In Table 3, the table title indicates that the analysis is for neck-shoulder pain. However, within the table, directly above Model 1, the phrase "NSP intensity" has been included. Could this phrase be removed and the table title modified (as needed) to clarify the associations that are being tested?

Authors: we have revised Table 3 according to the reviewer.

VERSION 4 – REVIEW

| | |
|------------------------|--|
| REVIEWER | Lisa Lix University of Manitoba, Canada |
| REVIEW RETURNED | 28-Jan-2019 |

| | |
|-------------------------|---|
| GENERAL COMMENTS | I have no further comments on the revised manuscript. |
|-------------------------|---|