PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A longitudinal observational study of back pain incidence, risk factors, and occupational physical activity in Swedish marine trainees
AUTHORS	Monnier, Andreas; Larsson, Helena; Nero, Håkan; Djupsjöbacka, Mats; Äng, Björn

VERSION 1 - REVIEW

REVIEWER	Chad Cook Duke University, United States
REVIEW RETURNED	29-Jul-2018

GENERAL COMMENTS	Thank you for the opportunity to review this paper. The study involved a high degree of effort and was associated with longitudinal evaluation of low back pain incidence or recurrence in in a very small sample of young marines, mostly men (N=53). The study involved weekly follow up and a unique HR method (repeated time to event) that I am not familiar with. The study is designed to identify risk characteristics for LBP incidence or recurrence. These characteristics are mostly well described in the paper, and more comprehensively in the tables. From what I read about the repeated time to event (Anderson Gill method) one needs a smaller sample, although I did not see a sample size estimate. Having performed standard HR designs, I am concerned that the sample is too small and I think this should be evaluated by a statistician. Some of the testing (e.g., Ranger tests) are not well described. I also was somewhat lost on how the accelerometer data were used since only 21 received this and data were calculated on only 5. Also, there is one week in which follow up was not possible and I'm not sure how this influences the modeling. My take away from this paper is that it fails to really state what is unique and important about this design. I know the HR method is novel, but do the findings provide anything of use for the military? Is the data transferable from this very unique sample of specialists? How does this study add to the literature and how does it impact future LBP? One last weakness that is identified by the authors is the definition of incidence. It may truly be recurrence and that's a very different thing. Having investigated this before the risk factors were unique between first time and recurrent LBP. My thanks and good luck.

REVIEWER	Prf. Dr. Leischik Roman University Witten Herdecke
REVIEW RETURNED	26-Sep-2018

GENERAL COMMENTS	This is a very low number of participants and without controls, The study have low essential information. For the futere you need more than 100 marines and 100 controls e.g. industry wokers/Police officers,

REVIEWER	Maria Kompoti Intensive Care Unit, Thriassion General Hospital of Eleusis, Athens, Greece
REVIEW RETURNED	16-Oct-2018

GENERAL COMMENTS	The authors have conducted a prospective observational cohort study aiming at evaluating the incidence and impact of low back pain in Swedish marines during their basic training course. It is a well designed study and data analysis has been carried out with implementation of an extended Cox regression model for recurrent events, as first described by Gill and Andersen (Annals Stat 1982;10:1100-1120). There are some points to be mentioned: a) I wonder what the effect of BMI in univariable and multivariable models was. Modeling individual body height without adjustment for body weight seems to me rather unreasonable. Moreover, I can see no rationale for categorizing body height as <1.80, 1.81-1.85 and >1.85, instead of fitting it as a continuous variable or as an ordinal variable using tertiles or quartiles. Nevertheless, I would like to see hazard ratios for BMI in univariable analysis (and if applicable, in multivariable analysis). b) It would also be interesting, perhaps in a future study, to implement fat-free mass index (using lipometry) as an independent predictor of LBP. c) The authors should discuss in Methods and Results sections the ways they assessed model fit and potential violations of proportionality assumption.

REVIEWER	Peter Smith University of Southampton, UK
REVIEW RETURNED	25-Oct-2018

GENERAL COMMENTS	The statistical analysis of this observational study of low back pain (LBP) in Swedish Armed Forces marines during their basic training course, presented in this paper, appears to be appropriate, although some clarification and further discussion are needed, particularly regarding the implications of the small cohort size. It is good to see some sensitivity analyses performed to assess robustness of the results. In the modelling, please confirm that time was measured in days. The Andersen-Gill model is a continuous time model, so might not

	be appropriate if weeks were used. In that case, a discrete time hazard model would be more appropriate. The results in Ingel and Jahn-Eimermacher (2014) suggest to me that your study with a cohort of only 53 marines is underpowered. As Vittinghoff and McCulloch (2006) state on page 717 “[Their] evaluation focuses primarily on confidence interval coverage for β1 and the related type I error rate of the test of H0 (β1 = 0), secondarily on bias in the estimate of β1, and only indirectly on variability and power.” Furthermore, they do not consider recurrent events. While the sample size is relatively small, the literature does suggests that the tests used in the analysis will have the correct type 1 error rates. However, I am concerned that the probability of type 2 errors is high. While you should not perform a post-hoc power analysis using the observed effects sizes, inspection of the confidence intervals for the non-significant parameters is insightful. This is the approach recommended by Vittinghoff and McCulloch (2006). Therefore, I suggest there is a discussion of the potential effects of any non-significant variables given the small cohort size. While confidence intervals are presented for the univariate analyses, presenting the results of the multiple variable regressions that include all the independent variables might aid this discussion. Please can you define what you mean by missing completely at random. I assume you mean what is sometimes called covariate dependent missingness and that you are controlling for all the relevant covariates (independent variables) in your (imputation) models. I think some discussion is required as to whether managing fewer than four pull-ups, physical training of fewer than three sessions per week and the other physical baseline tests are potential causes or consequences of (previous) LBP. While I understand that these are assumed to have been measured before the onset of the LBP episode, I am not convinced that ‘The presented results also highlight the need for regular physical training ...’ (my italics), although I believe that regular excises can reduce the incidence of LBP. Minor comments I would also provide the confidence intervals for the estimated incidence rates in the Abstract Results section. The superscript a on ‘(years)’ needs deleting in Table 3. The caption of Figure 3 contains some repetition and therefore should be shortened. Reference Ingel, K. and Jahn-Eimermacher, A. (2014) Sample-size calculation and reestimation for a semiparametric analysis of recurrent event data taking robust standard errors into account. Biometrical Journal, 56, 631–648 (DOI: 10.1002/bimj.201300090 631).

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Chad Cook

Institution and Country: Duke University, United States

Please state any competing interests or state ‘None declared’: I have none

Please leave your comments for the authors below

Reviewer’s comments ; Thank you for the opportunity to review this paper. The study involved a high degree of effort and was associated with longitudinal evaluation of low back pain incidence or recurrence in in a very small sample of young marines, mostly men (N=53).

The study involved weekly follow up and a unique HR method (repeated time to event) that I am not familiar with. The study is designed to identify risk characteristics for LBP incidence or recurrence. These characteristics are mostly well described in the paper, and more comprehensively in the tables. From what I read about the repeated time to event (Anderson Gill method) one needs a smaller sample, although I did not see a sample size estimate. Having performed standard HR designs, I am concerned that the sample is too small and I think this should be evaluated by a statistician.

Reply: Cautions has been taken to avoid over-fitting of our statistical models by only including potential risks based on empirical or scientific evidence. Furthermore, sensitivity analyses have been conducted, and model-fit / assumptions of all final models have been carefully examined before inference of the results. We therefore believe inference of the present results to be appropriate, not least because our sample consisted of 95% of the Swedish marine trainee population at that time. Potential concerns regarding our sample size have, however, now been further addressed in the article’s method section, see page 10 line 31, result section, see page 11 line 6, and the discussion section, see page 12 line 1.

Reviewer’s comments; Some of the testing (e.g., Ranger tests) are not well described.

Reply: The lower limb functional tests and The Ranger test were not analyzed in the present study, they were simply mentioned to provide clarity for the reader as they were conducted in the same protocol as the tests analyzed here. The Lower limb loading tests are composed of four parts; step-up (5 times/side), step-down (5 times/side), unilateral leg-rises from sitting (5 times/side) on a 40cm high bench, and 5 repeated deep squats. Perceived knee pain intensity is rated using the Borg CR 10 scale. For the Ranger test (lower-limb functional capacity test), the subjects perform 75 step ups on a 40 cm high bench with each leg, at a rate of 25 steps per minute. For this study, they wore a 12-kg standard military backpack. We have now revised the text on page 6, line 30-31, and we have also provided references for detailed test descriptions.

Reviewer’s comments: I also was somewhat lost on how the accelerometer data were used since only 21 received this and data were calculated on only 5. Also, there is one week in which follow up was not possible and I'm not sure how this influences the modeling.

Reply: Every week could not be monitored due to safety- and logistical reasons relevant in the marines’ work, and the collected data on occupational physical activity covered approximately 30% of the marine training course for 40% of its participant. As such, it is well over the recommended number of days to contras habitual physical activity in sedentary and light intensity. Being able to collect data for more weeks would of course have strengthened the analysis for the time spent taking part in vigorous activities. We do, however, believe that the data captured is sufficient to frame the

occupational exposures of the course and to aid the reader in the interpretation of identified risks, as the study intended. As data from objective monitoring was not used in incident calculations or regression analysis, the week where follow up was not possible did not affect the regression data, rather it diminished the weeks of the course covered with objective monitoring. In regard to the other analysis, the week lost to follow up was handled by incorporating it into the adjacent weeks, and adjusting the time variable to represent the duration for those weeks.

Reviewer's comments: My take away from this paper is that it fails to really state what is unique and important about this design. I know the HR method is novel, but do the findings provide anything of use for the military? Is the data transferable from this very unique sample of specialists? How does this study add to the literature and how does it impact future LBP?

Reply: While LBP in elite military personnel, such as marines, have been given increased international scientific attention in the last decade, this has most often focused on entry level training or deployment and not on their early career. Hence, knowledge from the subsequent early phases of military training, where preventive measures are likely to be very important for the reduction of future LBP, is lacking. The present study addresses this knowledge gap in regards to risks for LBP, and might aid future preventive actions in similar military units. We have now addressed this further in the background, see page 4 line 8, and hope that the rationale for this study is now made clear for the reader.

Reviewer's comments: One last weakness that is identified by the authors is the definition of incidence. It may truly be recurrence and that's a very different thing. Having investigated this before the risk factors were unique between first time and recurrent LBP. My thanks and good luck.

Reply: Based on the prevalence of back pain reported in our previous studies of active duty marines

(Monnier, A et al. Musculoskeletal pain and limitations in work ability in Swedish marines: a cross-sectional survey of prevalence and associated factors. BMJ Open. 2015;5(10):e007943.) and studies on the Swedish Armed Forces basic military training (Larsson H, et al.; [Scientific report (in Swedish); Rekrytering, urval och uppföljning]. Karolinska Institutet, Stockholm, 2013), we believe that the majority of the cases presented in the present study do not represent a true "first ever" back pain episode, but rather a new event in a course of recurrent back pain. We agree with the reviewer that our definition of a pain episode does not distinguish between a new "uniquely" first event and a symptom "flare up" associated with a previous event, and the present study can therefore not identify potential differences in underlying risk of "first ever" events or recurrent events. We have now addressed this in the discussion, see page 12 line 31. Thank you for your helpful comments.

Reviewer: 2

Reviewer Name: Prf. Dr. Leischik Roman

Institution and Country: University Witten Herdecke

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

Please leave your comments for the authors below

Reviewer's comments: This is a very low number of participants and without controls, The study have low essential information. For the futere you need more than 100 marines and 100 controls e.g. industry wokers/Police officers,

Reply: Yes, our cohort consists of a small but very homogeneous sample representing 95% of this specific population of Swedish marines, not a mix of different military (or police) positions that can

tend to dilute important and specific risk exposures for this group. While the loss of power due to the sample size could have been avoided by including the results from future yearly training courses (i.e. giving a larger sample), or by prolonging the follow up period to include time also after the course, we considered the homogeneity of work-related exposure from only one course to be more important for the present study's aims. However, limitations associated with our small sample are now outlined in more detail in the discussion, see page 12 and line 1. Thank you.

Reviewer: 3

Reviewer Name: Maria Kompoti

Institution and Country: Intensive Care Unit, Thriassion General Hospital of Eleusis, Athens, Greece

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

Please leave your comments for the authors below

Reviewer's comments: The authors have conducted a prospective observational cohort study aiming at evaluating the incidence and impact of low back pain in Swedish marines during their basic training course. It is a well designed study and data analysis has been carried out with implementation of an extended Cox regression model for recurrent events, as first described by Gill and Andersen (Annals Stat 1982;10:1100-1120). There are some points to be mentioned:

a) I wonder what the effect of BMI in univariable and multivariable models was. Modeling individual body height without adjustment for body weight seems to me rather unreasonable. Moreover, I can see no rationale for categorizing body height as <1.80, 1.81-1.85 and >1.85, instead of fitting it as a continuous variable or as an ordinal variable using tertiles or quartiles. Nevertheless, I would like to see hazard ratios for BMI in univariable analysis (and if applicable, in multivariable analysis).

Reply regarding BMI: While the use of the "standard" BMI categorization for overweight may not be valid for physically well-trained individuals (Mazic S et al. Overweight in trained subjects-are we looking at wrong numbers?(Body mass index compared with body fat percentage in estimating overweight in athletes.). Gen Physiol Biophys. 2009;28:200-4.),

such as marines, we agree with the reviewer that BMI could be seen as a measure of body composition that could affect the identified relationship between body height and back pain. On the basis of our previous results from this population, BMI was a priori considered to have a potential confounding association with body height in relation to back pain. However, our present analysis did not indicate that BMI constituted a necessary variable to control for, given that it did not change the hazard ratio for body height in the final regression models with >20%.

As requested, please see Table 1 below for BMI hazard ratios, p-values, and 95% confidence interval from uni-, final crude and final adjusted multivariable models. While significance level from univariate modeling would have allowed for inclusion in further model build, with the exception for the borderline significant final crude model for LBP limiting work ability, it does not reach significance ($p < 0.05$) level when included in the final adjusted models, and displays modest Hazard ratios. Furthermore, its inclusion in final models does not alter inference of body height's association to LBP in the present study.

Table 1. Inclusion of BMI in regression analyses of individual physical characteristics, work- and health-related risk variables: univariate, multiple final, and multiple adjusted hazard ratio (HR) for low back pain and low back pain limiting work ability during the marine training course.

Low back pain during the marine training course.

	Univariate			In; Final Crude Multivariable[a]			In; Final adjusted Multivariable[b]		
	HR	95% CI	p-value	95% CI	HR	p-value	HR	95% CI	pvalue
BMI	1.07	0.98-1.17	0.16	1.15	0.21	1.06	1.05	0.97-1.14	0.29

Low back pain limiting work ability during the marine training course.

	Univariate			In; Final Crude Multivariable[c]			In; Final adjusted Multivariable[d]		
	HR	95% CI	p-value	95% CI	HR	p-value	HR	95% CI	pvalue
BMI	1.11	0.95-1.30	0.19	0.97-1.33	0.12	1.19	1.00-1.41	0.049	1.13

[a]Included in model with; and Body Height ≤1.80 (m), Back Pain within 6 mo. prior to course start. [b]Included in model with; and Body Height ≤1.80 (m), Back Pain within 6 mo. prior to course start, adjusted for confounding effect of sex. [c]Included in model with; and Body Height ≤1.80 (m), Back Pain within 6 mo. prior to course start and Physical training ≤ 2 sessions/week [d]Included in model with; and Body Height ≤1.80 (m), Back Pain within 6 mo. prior to course start and Physical training ≤ 2 sessions/week, adjusted for confounding effect of sex and neck/shoulder pain previous to course start.

Reply regarding body height: The categorizing of body height as ≤1.80m, 1.81-1.85m (reference) and

≥1.86m was based on our previous results from this population (Monnier et al. Musculoskeletal pain and limitations in

work ability in Swedish marines: a cross-sectional survey of prevalence and associated factors. BMJ Open. 2015;5(10) and Monnier et al. Risk factors for back pain in marines; a prospective cohort study. BMC Musculoskelet Disord. 2016;17(1):319), and represents the tertiles of the SwAF marine population, which we hope is now clear to the reader, see Table 1, page 19 line 31. While we agree that continuous variables should be used when possible, our previous results indicated that at least a quadratic form, or preferably linear- or restricted cubic splines, would have been necessary to represent a curvilinear relationship of body heights association with LBP (also confirmed for present study by use of Martingale residuals). However, to avoid the risk of over fitting the present model, we elected to explore this association based on teritiles.

Reviewer's comments: b) It would also be interesting, perhaps in a future study, to implement fatfree mass index (using lipometry) as an independent predictor of LBP.

Reply: We do agree with the reviewer and will consider this for future studies in this population. Thank you.

Reviewer's comments: c) The authors should discuss in Methods and Results sections the ways they assessed model fit and potential violations of proportionality assumption.

Reply: Using methods such as tests based on reestimation described by Cleves et al. (Cleves M, et al. An

introduction to survival analysis using Stata.2010, 3ed ed., College Station: Stata press) interaction of analysis time with the independent variables and graphically through Schoenfeld residuals, the final models showed no violations of underlying assumptions of proportional hazards. Final models were also determined to have appropriate model fit as determined by generalized (Cox-Snell) residuals, Martingale residuals, deviance, DFBETA and likelihood displacement. We have now revised the text in the methods section, see page 10, line 31. Thank you.

Reviewer: 4

Reviewer Name: Peter Smith

Institution and Country: University of Southampton, UK

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

Please leave your comments for the authors below Please see the attached file.

Reviewer's comments: The statistical analysis of this observational study of low back pain (LBP) in Swedish Armed Forces marines during their basic training course, presented in this paper, appears to be appropriate, although some clarification and further discussion are needed, particularly regarding the implications of the small cohort size. It is good to see some sensitivity analyses performed to assess robustness of the results.

In the modelling, please confirm that time was measured in days. The Andersen-Gill model is a continuous time model, so might not be appropriate if weeks were used. In that case, a discrete time hazard model would be more appropriate.

Reply: For logistical reasons, to avoid recall bias and to avoid the risk to create a "false onset time" due to creeping onset of back pain, time was measured in weeks. Given that we have observed events on a discrete time scale (i.e. weeks) the risk of ties has increased compared to if the division of continuous time where conducted in finer units of discrete time (i.e. days). However, given the use of

a recurrent event model with discontinued risk, the number of ties of event times is still not to be considered large.

Together with the use the effron method to handle ties, we do not believe the division of continuous time in weeks has a significant effect on the presented models.

While we at this point cannot investigate if a finer division of continuous time would affect our analysis (i.e underestimation of true error variance), the model build was repeated for the main outcome and independent variables from block 1 using a negative binomial regression model, including exposure time. Given that no time-depending independent variables were included in the Andersen-Gill repeated timeto-event regression, a negative binomial model is expected to yield similar results, whilst not taking time to event in consideration (Jahn-Eimermacher A. Comparison of the Andersen–Gill model with Poisson and negative binomial regression on recurrent event data. Computational Statistics & Data Analysis. 2008;52(11):4989-97). This modeling confirmed the main result for the primary outcome; Previous back pain (incidence rate ratio 2.54, 95% CI 1.39-4.66) and shorter body height (incidence rate ratio 2.06, 95% CI 1.17-3.64) emerged as risk factors for LBP during the training course.

We do agree with the reviewer that a discrete time hazard model that allows intermediate censoring may have been more robust. Still, we elected the present model as it allowed the option to include recurrent events, time varying covariates (especially the ability to control for the potential confounding of previous pain episodes during the course), and “gap” time, i.e. not at risk when experiencing pain. We have discussed this extensively and believe the later to be fundamental when modeling recurrence in back pain in the military context.

Reviewer’s comments: The results in Ingel and Jahn-Eimermacher (2014) suggest to me that your study with a cohort of only 53 marines is underpowered. As Vittinghoff and McCulloch (2006) state on page 717 “[Their] evaluation focuses primarily on confidence interval coverage for β1 and the related type I error rate of the test of H0 (β1 = 0), secondarily on bias in the estimate of β1, and only indirectly on variability and power.” Furthermore, they do not consider recurrent events.

While the sample size is relatively small, the literature does suggests that the tests used in the analysis will have the correct type 1 error rates. However, I am concerned that the probability of type 2 errors is high. While you should not perform a post-hoc power analysis using the observed effects sizes, inspection of the confidence intervals for the non-significant parameters is insightful. This is the approach recommended by Vittinghoff and McCulloch (2006). Therefore, I suggest there is a discussion of the potential effects of any non-significant variables given the small cohort size. While confidence intervals are presented for the univariate analyses, presenting the results of the multiple variable regressions that include all the independent variables might aid this discussion.

Reply: We agree with the reviewer that the “sample” size is to be considered small (albeit very homogenous operating under the same employer with mainly the same work tasks), leading to a heightened risk for type II errors in general, and specifically the false exclusion of borderline significant independent variables. This is now further discussed on page 12, line 1-7.

We did inspect the confidence intervals for the non-significant parameters during our model build, as suggested by the reviewer, and made full use of the iterative “purposeful selection” process to inspect change in the confidence interval when including or omitting borderline significant independent variables, now described in the method section on page 10, line 6 and results section on page 11, line 6. However, this process did not indicate any borderline significant variable to have been wrongfully omitted. As suggested by the reviewer, however, the initial multiple regressions model is now included as an additional file, see supplementary Table 2.

Reviewer's comments: Please can you define what you mean by missing completely at random. I assume you mean what is sometimes called covariate dependent missingness and that you are controlling for all the relevant covariates (independent variables) in your (imputation) models.

Reply: To guide management of missing data, the reason and type of missing data, as well as assumptions of the missing mechanism were analyzed. Data were consider to be "missing completely at random" if the reason for data being missing was not dependent of the missing data itself (i.e. data on back pain not missing as a result of the participant experiencing back pain), nor predicted by the independent variables included the analysis. When the missing data could be predicted only by independent variables from planned analysis, the missing mechanism was considered to be "covariate

missing completely at random"(Vittinghoff E,et. al. Regression methods in biostatistics: linear, logistic, survival, and repeated measures models. 2005. New York: Springer). Based on these analyses, the data for outcomes and the DLL-ALE test were considered to be "missing completely at random" and the data on the kettlebells lift and the pullups tests were considered to be "covariate missing completely at random". We hope it is now clear to the reader, see page 8, line 34. Therefore, all relevant variables (planned to be used in the analysis and independent variables associated with the missing mechanism) were used to build the multiple imputation models.

Reviewer's comments: I think some discussion is required as to whether managing fewer than four pull-ups, physical training of fewer than three sessions per week and the other physical baseline tests are potential causes or consequences of (previous) LBP. While I understand that these are assumed to have been measured before the onset of the LBP episode, I am not convinced that 'The presented results also highlight the need for regular physical training ...' (my italics), although I believe that regular excises can reduce the incidence of LBP.

Reply: We do not have records of "first time ever onset of back pain" or back pain episodes earlier than six months prior to course start. As such, we were not able to identify if the exposure preceded the reduction in physical training or upper body strength further back in time than six months prior to course start. However, we could not identify any statistical evidence for an effect measure modification between back pain experienced within the 6month prior to course start and the amount of physical training or upper body strength, in relation to a new LBP episode, or LBP episode that affected the work ability, during the course. Still, physical training is recommended as preventive actions for back pain in both general populations and occupational settings. Taken together, we believe this highlights the potential role of physical training as preventive actions against future back pain episodes for marines displaying these identified risks. We have now added this discussion to page 13, line 29.

Minor comments

Reviewer's comments:I would also provide the confidence intervals for the estimated incidence rates in the Abstract Results section.

Reply: As suggested by the reviewer, the confidence intervals for the estimated incidence rates have been included in the abstract results section, see page 2, line 10-11.

Reviewer's comments: The superscript a on '(years)' needs deleting in Table 3. Reply: Corrected as suggested. Thank you.

Reviewer's comments: The caption of Figure 3 contains some repetition and therefore should be shortened.

Reply: The caption has now been shortened. Thank you.

VERSION 2 – REVIEW

REVIEWER	Chad Cook Duke University, United States
REVIEW RETURNED	05-Dec-2018

GENERAL COMMENTS	Thank you for your edits. My 10,000 foot take away is the following. 1. This is a small sample size of a distinct population. 2. The purpose statement is vague and the outcomes are less clear 3. The nature of the study is complicated as are the statistical methods. It is difficult to me to determine if the stats are correct for this sample size. 4. The paper is VERY detailed and long and I get somewhat lost from the purpose of the paper which is vague. 5. There are different methods and different sample sizes used for the outcomes. This further complicates the paper. 6. I think the paper is very well written, it is obvious the authors know what they are doing; it's just that the paper is very complex. It's a complex paper regarding a small sample size. I think we need more plain language around the purpose and a clearer explanation of the findings in the first part of the discussion

REVIEWER	Peter Smith University of Southampton, UK
REVIEW RETURNED	11-Jan-2019

GENERAL COMMENTS	While you have satisfactorily addressed my comments, given that time is measured in weeks, I suggest you also compare your final results with those from random effects logistic models for recurrent events. Time varying covariates can be included in these models, and a good overview is in https://www.bristol.ac.uk/media-library/sites/cmm/migrated/documents/recurrent-events.pdf. I also have a couple of minor comments: While retaining the information requested by the Editor, I think the title could be shortened to something like “A longitudinal observational study of lower back pain incidence, risk factors, and occupational physical activity in Swedish marine trainees”. Page 10, line 8: I suggest you clarify how you determine whether or not there has been an ‘incorrect omission’. For example, a variable with a relatively large estimated effect size and a wide confidence interval, resulting in it being non-significant, might be incorrectly omitted.

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Chad Cook

Institution and Country: Duke University, United States

Please state any competing interests or state ‘None declared’: None

Reviewer’s comments: Thank you for your edits. My 10,000 foot take away is the following.

1. This is a small sample size of a distinct population.

2. The purpose statement is vague and the outcomes are less clear

3. The nature of the study is complicated as are the statistical methods. It is difficult to me to determine if the stats are correct for this sample size.

4. The paper is VERY detailed and long and I get somewhat lost from the purpose of the paper which is vague.

5. There are different methods and different sample sizes used for the outcomes. This further complicates the paper.

6. I think the paper is very well written, it is obvious the authors know what they are doing; it's just that the paper is very complex. It's a complex paper regarding a small sample size.

I think we need more plain language around the purpose and a clearer explanation of the findings in the first part of the discussion

Reply: As suggested by the reviewer, we have now revised the text around the purpose, see page 4, line 38 to page 5, line 1, and believe that the rationale for the different study aims and methods is now clear to the reader. While we have structured our discussion in line with the recommendations of Smith, R, (Smith, R: The case for structuring the discussion of scientific papers. BMJ. 1999;318:1224–5) and as such strived to keep the summary of the principal findings very short, we have now revised this section, page 21, line 21-26, as requested by the reviewer. By setting the results in relation to the study aim, and revising the description of the principal findings, we hope that this summary section is now clear to the reader.

Reviewer: 4

Reviewer Name: Peter Smith

Institution and Country: University of Southampton, UK

Please state any competing interests or state ‘None declared’: None declared

Reviewer’s comments: While you have satisfactorily addressed my comments, given that time is measured in weeks, I suggest you also compare your final results with those from random effects logistic models for recurrent events. Time varying covariates can be included in these models, and a

good overview is in https://www.bristol.ac.uk/media-library/sites/cmm/migrated/documents/recurrentevents.pdf.

Reply: As suggested by the reviewer, we have now verified our final results regarding individual physical characteristics, work- and health-related risk variables, by analyzing the final multiple models for LBP and LBP limiting work ability also by random effects recurrent events logistic regressions, please see Table 1 below for OR ratios, p-values, and 95% confidence interval for final adjusted models.

Comparing our existing results with those that emerged from the random effects logistic model showed that the results are stable, i.e. our results from the Andersen-Gill repeated time-to-event regression can be confirmed. We therefore believe that inference of the results is appropriate in the present study. Thank you for your suggestion in this important passage.

Table 1. Random effects recurrent event logistic regression analyses of individual physical characteristics, work- and health-related risk variables: multiple final adjusted [a,b] odds ratio (OR) for low back pain (LBP) and LBP limiting work ability during the marine training course.

Variable	LBP Final Adjusted Multivariable[a]			LBP limiting work ability Final Adjusted Multivariable[b]		
	OR	**95% CI**	***p* value**	**OR**	**95% CI**	***p* value**
Physical characteristics						
Body Height ≤1.80 (m)	2.11	1.23	3.60	0.006	4.32	1.35 13.80 0.014
Rated health/health history						
Back Pain; within 6 mo. prior to course start	2.77	1.51	5.10	0.001	3.56	1.11 11.34 0.033
Physical training habits past 6months	Physical training; ≤ 2 sessions/week	2.47	1.03	5.94	0.044	

[a] Adjusted for confounding effect of sex

[b] Adjusted for confounding effect of sex and neck/shoulder pain previous to course start

Minor comments

I also have a couple of minor comments:

Reviewer's comments: While retaining the information requested by the Editor, I think the title could be shortened to something like "A longitudinal observational study of lower back pain incidence, risk factors, and occupational physical activity in Swedish marine trainees".

Reply: Changed as suggested.

Reviewer's comments: Page 10, line 8: I suggest you clarify how you determine whether or not there has been an 'incorrect omission'. For example, a variable with a relatively large estimated effect size and a wide confidence interval, resulting in it being non-significant, might be incorrectly omitted.

Reply: For omitted variables, we inspected the 95% CI, i.e. the lower limit, in relation to the size of effect estimate, which we have now clarified on page 15, line 40-41. For example, for a variable with relatively large estimated effect size, a lower confidence interval of 0.93 could potentially indicate that it had been falsely excluded. For a variable with a small estimated effect size, the same lower limit 95% CI would not raise the same suspicion. We did not, however, apply any a-priori considered cut-offs nor use any formal testing. Given the rather modest size of effect estimates in comparison to lower limits of the CIs, we do not suspect any false exclusion of potential risk factors.