Author's response to reviews

**Title:** Prediction model for unsuccessful return to work after hospital-based intervention in low back pain patients

**Authors:**

Ole K Jensen (olejesen@rm.dk)
Kristian Stengaard-Pedersen (stengaard@ki.au.dk)
Chris Jensen (chris.jensen@air.no)
Claus V Nielsen (clanil@rm.dk)

**Version:** 2 **Date:** 20 March 2013

**Author's response to reviews:** see over
**Reviewer:** Cecile Boot  
**Reviewer's report:**  
The authors have clearly put a lot of effort in this manuscript. Though I do find this an important topic, I have some major issues with this manuscript that need to be addressed.

**Major compulsory revisions**

1. Predictive versus etiological model  
This study aims to build a predictive model. This is conceptually different from an etiological model, where a specific risk factor is being associated with outcome, corrected for confounders. However, in the discussion section, it seems as if the authors interpret the predictors as if the aim of this study was etiological. For example, they seem to wonder why predictor X, known as a risk factor from literature, did not remain in the prediction model. However, this is not relevant, since in prediction models, statistical rules are applied, such that the variable with the (sometimes only slightly) stronger association remains, and another powerful predictor is left out.

*Answer:* We agree. In the revised manuscript, prognostic factors have been called risk factors in the first part of the manuscript, and later they are described as predictors.

2. Aim and organisation of the manuscript  
This is a very large manuscript, and because of the many tables, figures and long texts, I lost track. I feel that the authors should be more strict in selecting the information they have to present to answer their research question only. In the current manuscript, a lot of additional information (e.g., figure 1) can be left out. In line with this, I have doubts about the aim of this study, it is quite vague:  
- to identify risk factors  
- differences between patients with and without radiculopathy  
- validation of prediction model  
Identification of risk factors is not an issue here, as the authors state in the discussion paragraph, causal mechanisms cannot be revealed. Predictors of outcome is a better term.

*Answer:* We agree that the manuscript is large. However, as not only one, but two patient populations and two outcomes have to be described, it is difficult to shorten the manuscript into ‘normal size’.  
We have kept Figure 1 since the pain score and side-flexion are the only clinical variables included in the final model, and these two variables are fundamental for the definition of the three risk categories, and they may highlight the nature of low back pain as pointed out in the Discussion: ‘Restriction of side-flexion may stem from increased muscle stiffness.’  
The aims of the manuscript have been re-phrased, but we still consider it necessary to compare risk factors in patients with and without radiculopathy.
3. Definition of RTW
The researchers have chosen to look at RTW at 1 year follow up, rather than investigating the time until sustained RTW, which is more common. In the discussion section they argue that this is a more realistic measure than survival, since it also takes into account recurrences.
I do not agree with this. It is generally accepted that RTW for 4 weeks / 1 months is considered as successful RTW. One year is a long time, so the chance for new episodes of sick leave is quite likely.
In the current analyses, employees who had successful RTW at the start of the year, have been working successfully for 10 months, and two weeks before the follow up measurement got sick again, are now listed as unsuccessful RTW. I would strongly recommend to at least give insight in the patterns of sicklistings over the follow up year.
In fact, the outcome based on the way of measuring this is rather the chance of being sick listed (again or still) at follow up or not, rather that successful or unsuccessful RTW.
In addition to this, I have doubts whether the outcome is truly a measure for unsuccessful RTW. They rely on administrative data which has of course a huge advantage, with a 100% follow up rate! However, a disadvantage is that it is unclear whether participants are actually working. For example, those who are on job-training, further education, a supported job function are now in the group with unsuccessful RTW, whereas people who are unemployed are listed as successful RTW.

Answer: We have included RTW for 4 weeks during the year as a secondary outcome (initial RTW), but we have kept one-year RTW as the primary outcome, since we consider long term outcome in LBP patients as the most important outcome. We have changed it to one-year RTW for 4 weeks to facilitate comparison with initial RTW.
In all our publications, we have defined RTW as not receiving social transfer benefits, except from unemployment benefits. One may question this choice, but if we had chosen to register unemployed participants as not working at one year, another bias would be introduced: unemployed people may in fact be healthy and capable of working, if there was a job.

4. Radiculopathy as aim
The difference between patients with and without radiculopathy appears to me to be rather a finding than an aim beforehand. I would suggest to leave out this difference from the aim of this study, and include radiculopathy as one of the predictors. The part in the introduction (page 3, line 10 from below) can be left out. I would expect a study into differences between the groups with and without radiculopathy in more detail, but that does not happen.

Answer: We have omitted a potential difference between patients with and without radiculopathy as part of the aims, but we consider it necessary to keep radiculopathy as a variable and also to check for effect modification.

5. Work-related predictors
The authors claim that they have presented the work-related predictors in a
previous paper. However, when the aim is to make a prediction model, why not include all the information available? It is not of interest is work causes URTW, but if work can be a good predictor of outcome. I would suggest to include these in the analyses, and the prediction model might become better than it is now.

Answer: We have previously analysed work related risk factors (Stapelfeldt CM et al., 2011), but not with special focus on heavy physical work. Our data on this aspect are insufficient and have not been included. We have mentioned this in the ‘limitations’ in the Discussion.

6. Results
I do not agree with the authors that adding MRI to the model only relates to little changes. The OR in table 4 for income changes from 2.58 to 2.25, and for alcohol from 1.95 to 2.25, and for smoking from 1.62 to 1.83. This is more than 10% in all cases. I would recommend to include MRI as a predictor.

Answer: The important message in the previous manuscript was that the estimates were not weakened, rather strengthened, when the analyses were restricted to patients who had MRI performed. However, we have omitted that Table in the present manuscript.

We do not agree that it is a good idea only to include patients with MRI performed. In the beginning of the first study period, the patients had MRI on clinical indication, especially symptoms or signs of radiculopathy. Accordingly, only analyzing patients with MRI would induce severe bias.

7. RCT population
The source population are participants in an RCT. This group differs from the general population. This should be discussed in more detail.

In addition, the validation group is very small. I wonder if the paper would not improve without this validation step, since it does not add much.

Answer: When our study is compared to the study of Johathan Hill et al., 2011, we find our patients were even more disabled than the high risk group of Hill’s study as measured by the Roland Morris Questionnaire. So our patients are not average primary care LBP patients. This point has been added to the discussion.

A larger validation study group would be preferable, but we do not agree that the validation study group should be omitted, since this group was used for validation of the prediction model.

8. Ethical approval
Did all participants signed informed consent?

Answer: Yes, as written in the section of ethical approval.

9. Methods section
The analyses paragraph is not very clear. I do not know if this is a language issue, or an issue relating to statistical background. I would recommend to rewrite this section, and describe the analyses as different steps, one step at the time.

For example:
- when did the authors decide to include a variable into the multivariate model? P<0.10?
- did they use forward or backward selection?

Answer: All data have been re-analyzed and the ‘Analyses paragraph’ has been extensively revised. The statistical method is now described in detail making it possible for other researchers to repeat the study. The criteria for including and excluding variables in the multivariate analyses are described. Forward selection was used.

10. Discussion, page 15
No home ownership may be a proxy for level of income, or level of education. Again, this is a predictive model, so the authors are not looking into risk factors, but to factors to select high and low risk groups. Not drinking alcohol may be associated with having diseases or using medication that prohibits alcohol use. Another explanation may be that strong religious identity excludes alcohol. The group without alcohol is thus a selected group, and my hypothesis is that drinking alcohol does not have anything to do with increasing the chance for RTW!

Answer: Yes, we agree that no home ownership may be a proxy for low income, or, as we have described, a proxy marker for social vulnerability. In the previous model, it was adjusted for low income. In the new model, low income did not reach statistical significance to be included as a predictor in the multivariate model.
We agree that little use of alcohol use may reflect various aspects of life, and we have supplied the Discussion with some considerations. However, two other longitudinal studies have also identified little alcohol use as a risk factor, as referred to in the Discussion. Therefore, little alcohol use still may be a reliable predictor, although we do not know exactly why.

Minor essential revisions
Title
The first part states unsuccessful RTW, the second part states good prognosis. This is a contradiction.

Answer: We agree that the previous title was not optimal. The title has been changed.

Abstract, methods
- It is unclear what is meant by ‘117 other similar patients’. Similar to what/who?
- The analyses are stated in the results section, but should be listed in the methods section

Answer: We have re-phrased the sentence ‘117 other similar patients’, and we have reorganised the Methods and Results section.

Results
- first sentence, of ‘Clinical baseline etc.’: I do not understand what the authors mean by corresponding to 6.3 of 10?

Answer: We have supplied the text with explanation. Pain VAS scales are usually defined as a 0-10 scale.
- The strongest associations with U-RTW were demonstrated by ....: I do not understand why total pain score and side flexion are considered the strongest associations; where is that based on? Which data, and are they presented in Table 2?

Answer: When following the statistical procedure as described in the Analysis section, the pain score and side-flexion were the only two clinical variables left in the model (except effect modification variables). The strengths of the associations regarding these two variables are illustrated by Table 2 and Figure 1.

Please avoid sentences describing a non-significant trend. Not significant implies no association.

Answer: We have avoided the expression tendency.

- Health is listed in table 2 as a clinical variable, but in the text it is described as psychosocial / lifestyle variable. Please change in Table 2.

Answer: We agree, and Table 2 has been changed accordingly.

Please stick to one sequence. The sequence in the text is different from the Table, which makes it a challenge to find the numbers in the Table to match the text.

Answer: The text has been revised extensively, and we believe that it is now easier to read the text and tables.

- Is an increase in AUC from 0.80 to 0.84, and an increase in % classified correct from 74% to 76% relevant?

Answer: Yes, it may be relevant. Adding one statistically significant variable to the model when AUC is high only increases AUC a little.

- Page 10, non-specific LBP vs. radiculopathy: it appears that the authors have looked into subgroup analyses first, after which they included an interaction term, which turned out to be not significant. However, it is more common to include an interaction term first, and when significant interaction is present, a subgroup analysis is justified. This is another way to reduce the size of this manuscript!

Answer: We agree, and we also analyzed for interaction at the beginning, as described. However, the interaction variables should be included last in the multivariate model. The Analyses section has been revised extensively and should now be more comprehensive.

Referee 2:

Reviewer: Vicki Kristman

This is a clear, well-written manuscript with a well defined research question: to
determine the association between unsuccessful return to work (U-RTW) and baseline clinical, psychosocial, and lifestyle characteristics; to compare those factors in patients with and without radiculopathy; and to develop and validate a prediction model for U-RTW. A cohort study was developed using baseline data from a randomized controlled trial and linking to social transfer payments to determine RTW outcome one year later. This resulted in no lost to follow-up as all participants were able to be identified in the database. This is a significant strength of this study.

No comments

1. Introduction, pg 3: Previous prediction rules examining RTW after low back pain should be introduced with a discussion as to what this study adds beyond these. Two of these off the top of my head are “Dionne C et al. A clinical return-to-work rule for patients with back pain CMAJ 2005”, and “Heymans MW et al. Return to work in a cohort of low back pain patients: development and validation of a clinical prediction rule”.

Answer: We have described what this study adds. However, we have not referred to the studies mentioned above, as the first paper describes a prediction model in primary care, and the second paper describes risk factors that were not validated in another population.

2. Materials and Methods, pg 5, paragraph 2: all patients included in the study were referred by a GP. There is a high likelihood of a referral bias leading to selection bias in this study. The nature of this bias should be discussed and highlighted as a limitation in the discussion section.

Answer: We agree, and we have supplied with a discussion of this aspect in the Discussion section. When looking at patients in secondary health care, bias due to referral from GPs may not be disadvantageous, but rather may reflect usual care.

3. Materials and Methods, pg 5, last paragraph (Inclusion criteria): LBP had to be the prime reason for the sick-listing and at least as bothersome as any possible pain elsewhere – this suggests to me that patients with fairly significant comorbidities could have been included in the study. Yet, there is no accounting for these comorbidities (for example; diabetes, arthritis, coronary conditions, etc.) in the analysis. Uncontrolled confounding influences due to these comorbidities may bias the results. If no information on comorbidities is available to include in the analysis, then at the very least, this potential source of confounding needs to be documented in the discussion.

Answer: LBP patients often have pain elsewhere as for instance in the neck. This specification was necessary to avoid patients with for instance neck pain as their major complaint. Self reported comorbidity was reported in the questionnaire, but was not included in the analyses, as we considered data insufficient. However, less than 2% reported diabetes, only one patient reported cardiovascular disease and very few patients reported arthritis. Therefore, we do not believe comorbidity to be important in the present study. Furthermore, comorbidity was not among the documented risk factors in a previous systematic review (Hayden JA et al., 2009).
4. Materials and Methods, pg 8, Statistics, paragraph 5: The use of a median cut point for dichotomizing pain and flexion scores is not that useful clinically. A validated cut-point that would be the same across samples would be more useful.

Answer: Using the medians is an objective way to analyse data, as they are not influenced by the association with outcome. To our knowledge, validated cut points of the pain scale used and side-flexion do not exist.

5. Results, pg 11, Prediction model: The 7 psychosocial and lifestyle factors were combined to one variable with 7 categories assuming that they contributed equally to the risk for U-RTW. However, we know from Table 4 that this is a faulty assumption. The strengths of the associations for the 7 variables range from 1.83 to 3.13. How do the authors justify this approach when the assumption they base it on is incorrect?

Answer: In the new model only four dichotomous risk factors were included in the common ‘4 risk factor variable’, but the new model may be criticized similarly. The answer will be: as noted in the discussion, a difference in Odds Ratio of 1.93 vs. 3.75 only represents a minor difference in RTW of 22% vs. 30%. Therefore, it only results in a minor approximation.

6. Discussion, pg 14, paragraph 3: Why do the authors consider the one-year RTW to be a more realistic measure of RTW than survival analysis of registered continuous work for a four week period? Yes, RTW-status at one year may account for patients sick-listed again after having returned to work for a period; but, one does not know if the reason for the subsequent sick-listing is due to the LBP or not. This highlights an important limitation of the social transfer payment data that should be mentioned in the limitations section: there is no diagnosis code associated with Danish social transfer data. Due to this important limitation, survival analysis would be a better approach for two reasons: 1) It is more likely that that reason for being sick listed initially is due to the LBP than at one year later, and 2) survival analysis models time to the event of RTW. Given LBP and the associated sickness absence is so recurrent in nature, it could just be due to chance whether or not a particular individual happens to be on or off work at a point one year later. Therefore, understanding the time to RTW is more useful than knowing whether or not a person is at work on a particular day (especially when you have the cohort data available (i.e., time to event data being used in a cross-sectional manner is statistically wasteful). If recurrences are of interest, these can also be modelled using survival analysis – the analysis is much stronger than a logistic regression approach.

Answer: We have supplied with initial RTW as a secondary outcome which answers many of the questions posed above. These data have been analysed by survival analyses previously (Jensen C et al., 2011 and Stapelfeldt CM et al., 2011). However, the survival analysis also required information on initial RTW from the Danish social transfer register which does not include diagnosis codes. When looking at initial RTW, this may be no great problem, as we believe that the risk for sick-listing due to another diagnosis during improvement from low back pain is probably small. It may
be a greater problem in the one-year registration. This aspect is discussed in the Discussion section. However, the validity of our data is supported by the close association between initial and one-year RTW.

Minor Essential Revisions
1. Introduction, pg 3, paragraph 3: Reference 5 has been cited twice at the end of the first paragraph

Answer: Apologize, the fault has been corrected.

2. Materials and Methods, pg 6, Baseline variables, second paragraph: What is considered a red flag for the purposes of this study? Please define.

Answer: We have used the term ‘red flag’ as described in the literature: age of debut <20 years or >55 years, fever or unexplained weight loss, progressive unrelentless pain, history of cancer, use of steroids or HIV, thoracic pain, widespread neurologic signs or structural deformity. We have not defined this term in the text, as we consider it basic knowledge (Waddell G. The back pain revolution, 2004).

3. References, pg 20 & 21: References #4 and #19 refer to the same citation.

Answer: Apologize, has been corrected.

Referee 3:

Reviewer: Sheilah Hogg-Johnson

Trying to identify a prediction rule for RTW outcomes for LBP has merit. However, there are a number of issues with both the conduct of this study and the reporting of it that require attention.
1. Is the question posed by the authors well defined? Yes

2. Are the methods appropriate and well described? The methods are not well enough described, and it is difficult to tell how appropriate they are.

Answer: The methods are now so well described that other researchers may be able to repeat the study. We also think that the methods are appropriate for establishing and validating a prediction model.

3. Are the data sound? It is difficult to assess if the data are sound. The description of the measures is not well presented.

Answer: We have refined the description of the measures. However, some of the measures are Danish and not internationally well-known, although they are validated and some of them also translated, as for instance the Common Mental Disorder Questionnaire (described in the Discussion section).
4. Does the manuscript adhere to the relevant standards for reporting and data deposition? The figure has some issues, identified below, the tables are difficult to read and have things missing in them (see below)

Answer: Please, see below.

5. Are the discussion and conclusions well balanced and adequately supported by the data? No

Answer: All data have been re-analysed and the reporting of data in the Results section as well as the Discussion section is totally different from the previous version of the article.

9. Is the writing acceptable? Could use a good edit for grammar, words missing in places, order of material in the methods section.

Answer: The writing has been revised again, and the Methods section has been reorganised.

1. The manuscript is difficult to read and follow and contradictory in places.

Answer: We believe that the revision has made it easier to read and understand text and tables.

2. The introduction mentions some literature that has considered prognostic factors or prediction rules for LBP but it is missing a lot of relevant literature. (see reference list provided at end) A justification for this study is that there are no validated prediction models for secondary care – by that, I assume this means for LBP cases past the acute phase. I think there is some relevant literature, though – see the reference list.

Answer: We have not been able to track the references, as no reference list was presented at the end. However, we have checked the literature again, and we only found inception cohort studies, but no study presenting a prediction model that may be used in sick-listed LBP patients referred to and managed in secondary health care.

3. The methods are confusing and not adequately described. The description of the patients has to be very clear here. The subjects were originally recruited for two different randomized studies (although apparently recruitment for the two was identical). For the intervention studies, the main issue for validity is internal validity and so generalizability is of less interest. However, here, for this prediction rule purpose, generalizability and representativeness become very important. There is insufficient description of the recruitment process to understand this. How were GPs recruited? How many of those approached agreed to participate? How did participant GPs compare to non-participants? What instructions were GPs give regarding recruitment of patients? Did they log who they approached, who accepted and who refused? Were all potentially eligible patients approached? How did participant patients compare to eligible non-participant patients?
Answer: The subjects were not recruited for two different studies, but were recruited in two study
periods. The validation study followed the original study, and GPs were informed that we had
finished the first study period, but that we still recruited and treated patients in the same way as
previously.

GPs were recruited by letters and meetings. There was no claim towards GPs of agreement in
participation. We have not analyzed the differences between GPs referring and not referring
patients. GPs were informed about the purpose of the project as well as inclusion and exclusion
criteria by letters and website. They informed their patients accordingly. No, the GPs did not
register who of their patients accepted or refused. No, we do not assume that all eligible patients
were approached, rather we have data on sick-listing in general indicating that only a percentage
of eligible patients were recruited (15-20%?). We will not be able to provide data comparing
eligible non-participants with participants. However, we believe that the patients are representative
for patients referred from GPs to secondary health care, and we have provided arguments for this
point of view in the Discussion section.

4. A lot of detail is given about the interventions, and yet this is not the focus of
this paper. Do the interventions matter here? Were they effective? Did they alter
the course of LBP and change U-RTW? If so, they should be included as factors
in the study. If not, then I think there is far too much detail given about them here
and there should just be a reference to the publication that describes the
effectiveness (or lack) of those interventions.

Answer: We have shortened this part of the Methods section. The results of the randomised study as
well as prognostic factors for pain and function have been published elsewhere, and these
references are also mentioned in the text (Jensen C et al., 2011, Jensen OK et al., 2010, Stapelfeldt
CM et al., 2011).

5. Second paragraph on page 6 is confusing. Inclusion/Exclusion suggest that
sick-listed from 4-12 weeks is inclusion yet there are subjects in the study outside
the range. An explanation is provided, but then the inclusion and exclusion as
given are not correct. Likewise, indications for surgery given as exclusion and yet
samples include people who had surgery. Either it is an inclusion/exclusion or it
is not – which?

Answer: We have explained why a few patients were included in spite of minor inaccuracies
regarding length of sick-listing, and these inaccuracies have been accepted by the reviewers and
editors in previous publications. We do not believe this aspect to have any influence on outcome.
About 10% were operated because of lack of improvement by conservative care which is usual
practice when managing low back patients. Planned surgery before inclusion was something
different. It was an exclusion criteria, because it would make it difficult to implement the
interventions.

6. Table 1 and Table 2 should have row to row correspondence – same order
would be helpful too. So if the 4th row in table 2 is the LBP classification into
non-specific LBP w/o radiation, LBP with pain below knee and/or radiculopathy,
then the 4th row of table 1 should describe how that classification was defined
and what measurements were used to make it.
Suggested columns for table 1:
i. Variable name
ii. Source (e.g., questionnaire, clinical exam, imaging, combination of questionnaire and imaging)
iii. Definition and/or Categories and/or Questionnaire Items Used and/or Scale
iv. Whether classified as Clinical or Psychosocial/Lifestyle (this is distinguished in analysis, so reader should be told which are which)
v. Ranking as to whether well-known risk factor, suspected risk factor or potential risk factor (these are distinguished in analysis, so reader should be told which are which)

Answer: We have chosen not to follow the proposal as suggested above, because the logic of the two tables is different. Instead we have marked well-known risk factors by 'w' and potential risk factors by 'p' hopefully making it more clear, in which order the variables were included.

7. An explanation of what a “box scale” is would be helpful.

Answer: Box scale is the same as ‘Numeric Rating Scale’. The expression has been changed.

8. Be consistent and accurate with how information is provided in table 2. For instance, for Sex, the left hand column indicates female/all, but what is actually provided is female/male. Sometimes “n (%)” is provided and sometimes just “%” – I recommend always providing both n and % for categories or binary.

Answer: We apologise, has been corrected.

9. Last paragraph on page 6 describes process for MRI in detail. MRI results are used to create one of the variables used in analysis – but not everyone got MRI. How was the variable created if there was no MRI and how does that impact the measurement properties of the variable?

Answer: The diagnosis ‘disc herniation without radiculopathy’ was only possible in patients who had MRI without having radiculopathy clinically. After changing outcome to one-year U-RTW for 4 weeks this diagnosis was not associated with outcome and the variable did not qualify for inclusion in the multivariate model. In any case, the number of patients with this diagnosis was too small to allow for a final conclusion. The criteria for performing MRI are described in the Methods section.

10. P 7 Follow-up variables: does this description tell us that anyone with sick-listing, job-training, further education, supported job-function or disability pension in the week leading up to the one-year date would be classified as U-RTW? This is not clear.

Answer: Yes, the text has been clarified.

11. If the results from the two intervention studies have been published with corresponding descriptions of the recruitment process, they should be cited somewhere in the methods. If not published, why not?
Answer: Only one intervention study has been published, and this study is cited in the methods (Jensen C et al., 2011). A subgroup analysis has been made and tested in the validation study group (cited as ref. 51, Stapelfeldt CM et al).

12. I am surprised that the investigators were able to run two intervention studies, with randomization without getting Research Ethics Board Approval. This would not be allowed in my own jurisdiction. Can they demonstrate that they comply with the Declaration of Helsinki?

Answer: Yes, all patients signed informed consent. We attach a paper from the regional research ethics committee.

13. Page 8 Statistics - There is a description of how reliability of classification is assessed without describing the classification before this point in the manuscript.

Answer: It is described in Table 1, and a reference note has been added.

14. As per suggestions for tables 1 and 2 above, if different variables are treated differently in the analysis, then more detail has to be provided about which variables are which, why they are classified differently and how the variable selection went on.

Answer: We have clarified this, please see answer to question 6.

15. There is not enough detail of the variable selection process here for the process to be repeatable. How was decision made to include or not include a variable in a model?

Answer: The statistical analyses are now thoroughly described in the Analyses section, including the decision process in regard to including and excluding variables in the multivariate analyses. We believe that the description of the analyses is now sufficient precise to allow other researchers to repeat the study.

16. How was collinearity checked and what was done about it when detected?

Answer: The process of handling collinearity is now described in detail in the Analyses section.

18. Why were clinical factors separated from psychosocial/lifestyle factors and why were psychosocial factors lumped with lifestyle factors? Why are questions about work so limited when the situation at work has been prominent in other prognostic studies for RTW for LBP?

Answer: As mentioned in the discussion, the clinical variables were analysed separately because of the long lasting discussion of the importance of clinical vs. psychosocial risk factors. Psychosocial and life style risk factors were analysed together because they may be interrelated, as for instance smoking and social status. Work related factors have been analysed separately and have also been published, but focus was not on physical job demands (Stapelfeldt CM et al., 2011). Accordingly, our study does not identify
the role of physical job demands as compared to the predictors we identified. In the discussion we have described the limitations for using the model in patients with heavy physical work.

19. How many different models were fit? I see one with clinical factors, one with psychosocial/lifestyle factors, one combining clinical and psychosocial/lifestyle factors and then later on page 8 one that includes pain/side flexion and psychosocial/lifestyle factors. Why so many?

Answer: The Results section has been extensively revised. It should now appear more clearly, how many models were established, and how the analytic process resulted in the final model. We have added an Appendix with a more detailed description of the models.

20. Why was 5% chosen for significance? Was the sample size selected with consideration for power, detectable change, number of variables considered?

Answer: A 5% level for statistical significance was chosen because this is the usual level. The sample size was given beforehand and was based on power calculation indicating the minimum number of participants required for the randomised intervention study (Jensen C et al., 2011).

21. Why dichotomize and then count risk factors? Could there be merit in looking at gradients over, say, level of pain?

Answer: Some of the psychosocial and lifestyle risk factors were dichotomised in the questionnaire, some were ordered and some were non-ordered categorical. If they were statistically significant in the univariate analysis, some of the categorical variables were dichotomised when used in the multivariate analysis to make the multivariate model more simple, and to create more room for other variables.

22. AUC 0.79 is modest.

Answer: We agree, but it may be useful. A high AUC reflects to some degree the diagnostic power of the model (to predict U-RTW), but it may only be partly related to the crucial question: Is the model able to predict RTW in the following cohort (the validation study group).

23. The final model identified does not have a very good accuracy, particularly in the validation sample and therefore, shouldn’t be recommended for use. Also, it is not clear how the final model is meant to be used practically. If developing a prediction rule such as this, I would expect some attention paid to the different kinds of mistakes one might make if applying the rule and the implications of those mistakes (e.g., false positive or false negative).

Answer: We agree, but re-analyses of the data have now resulted in a better model with better ability in predicting outcome in the validation study group. In addition, we have provided tables showing how the model may be used in daily practice. This improvement was caused by inclusion of the ‘bodily distress’ variable in the model which was not included in the previous model due to insufficient handling of collinearity.

24. Figure 1 – I would expect the solid thicker line to be the same in the top and
bottom graphs (e.g., the two graphs for total pain score and then again for the two graphs for side-flexion), but they aren't. Why are the scales on the horizontal axes in these pairs of graphs different? In black and white, the reader will not be able to distinguish which line refers to radiculopathy and which to non-specific low back pain.

Answer: The scales were different because of the smaller number of patients with radiculopathy. The X-scales have now been adjusted and are similar in all four sub-figures, and the total figure is now made in a way that makes it possible to read it in black and white.