PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A new stratified risk assessment tool for whiplash injuries predicts recovery
AUTHORS	Kasch, Helge; Kongsted, Alice; Qerama, Erisela; Bach, Flemming; Bendix, Tom; Jensen, Troels

VERSION 1 - REVIEW

REVIEWER	Prof. Dr. Michele Curatolo Head, Division of Pain Therapy University Department of Anaesthesiology and Pain Therapy University of Bern, Inselspital 3010 Bern, Switzerland I have no competing interests.
REVIEW RETURNED	11-Oct-2012

GENERAL COMMENTS	Kasch et al performed a prospective prognostic study in whiplash. They used criteria from an early study in a subsequent large cohort of patients. Such studies are very difficult to perform and there is definitely a need for them. The results seem to identify criteria that can help detecting patients in the acute phase of a whiplash injury who are likely to develop chronic pain. In my opinion, the main relevance of the findings consists in the possibility to better select patients in the acute phase who would qualify for studies on preventive and treatment strategies. In my opinion, at this stage the data cannot be used for clinical decision making. Together with positive aspects, the study has some limitations that I list below. The paper is hard to read, as the presentation is unclear and incomplete at different places of the manuscript. The therapeutic interventions are mentioned pretty briefly under “study population”. This issue should be presented separately and expanded. The description of the treatments is confusing. Initially, three parallel treatment groups are mentioned; then, it seems that two treatments have been applied to the low risk group. Please re-write and clarify. I suggest that the treatments be introduced in the flowchart. The flowchart divides patients into two groups (low and high risk), but it is not clear what actually differentiates these two groups, since the right and left arm look almost identical. Please clarify.

	The palpation test that you used for the “clinical assessment” is of unclear origin. The results of this examination may be very examiner-dependent. Are there any data on its inter-rater reliability? Primary outcome: please specify whether all the three criteria or only one have to be fulfilled in order to categorize the outcome as “handicap”. The statistical analysis is not explained, which is unusual for such a complex study. What kind of “non-parametric statistics” has been performed? What methods have been used to analyze the influence of the different predictors on the outcome? How did you analyze the effect of the three treatments and their potential confounding influence? After I have read the results, it seems that you compared the different strata for the different parameters. This can certainly be done, but prognostic studies are typically analyzed by more complex procedures, e.g. multivariate analyses. Please comment. At the end of the results, pressure algometry is reported, bus this procedure is not presented in the study aims and is not described in the methods section. It would be relevant to know more on the clinical performance of the stratification that you used. This is typically done by computing sensitivity, specificity and likelihood ratios of the tests. At the end of the discussion you mention problems related to categorizing patients as being at high risk. The message is obscure to me, please clarify. In this respect, see my comment above: as long as the predictive value is not quantified, it is hard to define the role of the stratification for clinical decision making. It can still be said that knowledge on the prognostic factors can help selecting patients for studies on preventive strategies, since it makes more sense to enrol patients who are more likely to develop chronic pain. I suggest that you stress this point. MINOR ISSUES I do not understand what “segregated” means, see e.g. the sentence in the abstract: “Bio-psychosocial factors were significantly segregated from the first assessment by risk strata”. Please check if this is a correct english term. Under key messages, please delete the term “more or less” and specify what you mean by “out of work”. Under “strengths”, the third item would actually fit under “limitations”, since the need for further validating the score is not a strength. Please remove “more or less” from the 1st sentence of the introduction. Exclusion criteria: I guess that the “significant past pain conditions” were exclusion criteria, which is not clearly specified. Also, neck pain and headache is mentioned under the bullet list and again few lines

	below, this time specifying the VAS. Please correct and delete the repetition. Please explain the abbreviation CROM. The CROM scale indicates that the higher the score, the worse the points. Please explain the score for those who are unfamiliar with it.

REVIEWER	Samuel McLean, MD, MPH Vice Chair, Research, Department of Anesthesiology Attending Physician, Department of Emergency Medicine University of North Carolina, Medical School Wing C, Chapel Hill, NC, USA
REVIEW RETURNED	21-Oct-2012

GENERAL COMMENTS	This is an excellent and important study.

REVIEWER	Dr Alison Rushton Senior Lecturer Physiotherapy / Physiotherapy Research Lead University of Birmingham United Kingdom I have no competing interests with this study.
REVIEW RETURNED	24-Oct-2012

REPORTING & ETHICS	Although CONSORT is not appropropriate as the study is not reporting a trial but a study embedded within a trial.
GENERAL COMMENTS	There is valuable data here to inform assessment and management of patients following whiplash injury that is important to publish. I would urge the authors to consider the feedback to improve clarity of the study and its clinical messages. Thank you for the opportunity to review this interesting and large prospective study. There is valuable data here to inform assessment and management of patients following whiplash injury that is important to publish. However the article does require considerable attention before it is acceptable for publication, as clarity of methods and results are difficult for the reader to access in its present form. In particular, further discrimination of content will assist clarity of the results and key clinical messages. Abstract 1. The abstract contains abbreviations that affect clarity. It would be best to use all words in full for the abstract. 2. The results need to be more clearly reported in relation to the defined outcomes of interest. The results need to accurately reflect the main text. 3. There is no analysis / discussion of findings within the abstract. 4. The conclusion needs to relate specifically to the defined

	outcomes of interest. The point re bio-psychosocial issues is unclear. Article summary 5. This summary needs to reflect accurately the content and terminology of the finalised article. 6. The detail of the RCT interventions in the limitations section is not relevant. It would be clearer to keep the focus to this study. Introduction 7. Is the premise that the Spitzer WAD classification was designed to predict outcome accurate? This is also mentioned in the conclusion. 8. What is the quality of the existing literature in this area? 9. The previous study (Kasch et al, 2001) and the development of the tool to assess risk merits further consideration in this section to inform the reader. 10. It is unclear why the other factors identified in the literature were not included in this study (first paragraph P11). This appears to be due to the timing of this study with data collection preceding this later work. The more recent literature is therefore best included in the discussion for evaluating this study's findings. Materials and methods 11. The design of this study within the RCT merits further consideration. Are there any implications of this design (multiple interventions across two trials) e.g. any potential treatment effects for your conclusions? Beyond a couple of brief mentions this point is not addressed. 12. This is not reporting a trial and therefore the Consort checklist is not appropriate. 13. The clarity of this study as distinct to the trials needs to be clearer throughout. Much of the content re the trials can be removed as it is not relevant to this study. At present the inclusion of trial information is confusing for the reader. 14. P12 refers to a group of 'low risk patients' (line 31) and this is unclear. In looking back at the previous trial it refers to allocation of low risk patients to this trial (Kongsted et al, 2008) following an allocation scoring system. The high risk patients were allocated to a different trial (Kongsted et al, 2007). Please clarify and discriminate this content so that the reader is clear. For example can this be clarified within the inclusion criteria? Are you referring to participants of both previous trials being included in this study? If referring to participants from both trials is this risk categorisation relevant to the current study? 15. It is unclear why such a broad range of WAD grades were included (I-III) (P13). This requires justification. Were WAD classification 0 patients excluded?

	16. The exclusion of significant headache or neck pain is unclear (P13). The rationale and detail of this requires explanation. Does the later point (P14, line 22) link to this criterion? If so why were these patients excluded? 17. What is meant by 'significant past pain conditions were in detail:' and the list afterwards? (P13/14) 18. The validity of the total risk assessment score merits mention and weighting of individual components is not mentioned (P15). Are the included ROC curves from the previous study or current data? This requires explanation and if from the previous study a reference rather than reproduction of the curves would suffice. 19. There are differences in the description of outcome measures used throughout the paper which is confusing for the reader. The terminology and clarity of outcome measures needs to be clear throughout. The outcomes of interest at 1 year set a priori are detailed as: handicap, NRS neck pain, NRS headache, and Copenhagen disability (pages 18/19) but this is different to the abstract and is confused by the clinical assessment content and results sections where other outcomes are reported. Can the clinical assessment section be deleted (does not appear relevant to this study?) and the results section focused to the outcome measures of interest? 20. The distinction of primary and secondary outcomes is relevant to the trial but not this observational study of risk factors. The follow up dates of 3 and 6 months from the trial are not relevant to this study as it focuses on 1 year follow up. Results 21. The number of participants included needs to be clear on P21. 22. The risk categorisation into low and high risk is again mentioned here (P21) and in the flow diagram and appears irrelevant in this study – see point 15. The results of the previous trial are not relevant to the results of this study. 23. No detail is provided of the risk stratification in this section and the numbers in the strata. 24. This section needs to relate specifically to the a priori defined outcomes e.g. why is handicap not reported as on P18? New outcomes are introduced in this section e.g. EIS, assessment of physical job demands, pressure algometry etc. Some outcomes are not reported e.g. NRS neck pain and Copenhagen disability. This section needs to be considered further to avoid the potential of selective outcome reporting / data mining. Discussion and conclusions 25. There is no discussion of the findings in the context of the existing literature?

	26. The point on P24 lines 37-45 is unclear. A lot of time is spent discussing the previous study on the risk assessment score that would be best in the introduction. 27. The points re division of patients into two groups (P25) are unclear as the patients were not in two groups for the analysis and presentation of results. Is this point about potential treatment effects and best included under 'limitations'? 28. Introduction of bio-psychosocial factors at this discussion stage is unclear as it is not mentioned earlier in the paper. The conclusion re bio-psychosocial is also unclear. 29. Sample size for this study and number of participants in the strata are not mentioned. 30. The exclusion of patients not working prior to the injury for the main outcome of interest needs to be acknowledged as a key limitation. How many participants were therefore excluded from the analysis? 31. How are findings from the literature consistent with this study's results? (P27 lines 27-34) 32. The clinical messages from the paper can be clearer to assist the reader. Issues of presentation 33. The writing style, grammar and meaning are not always clear. E.g. 'suffering more or less' (P9). 34. Investigators should be 'blinded' rather than blind-folded (P19).

VERSION 1 – AUTHOR RESPONSE

Reviewer: Prof. Dr. Michele Curatolo
Head, Division of Pain Therapy
University Department of Anaesthesiology and Pain Therapy University of Bern, Inselspital
3010 Bern, Switzerland
I have no competing interests.
Kasch et al performed a prospective prognostic study in whiplash. They used criteria from an early study in
a subsequent large cohort of patients. Such studies are very difficult to perform and there is definitely a
need for them. The results seem to identify criteria that can help detecting patients in the acute phase of a
whiplash injury who are likely to develop chronic pain. In my opinion, the main relevance of the findings
consists in the possibility to better select patients in the acute phase who would qualify for studies on preventive and treatment strategies. In my opinion, at this stage the data cannot be used for clinical decision making.
Together with positive aspects, the study has some limitations that I list below.
a. The paper is hard to read, as the presentation is unclear and incomplete at different places
of the manuscript.
i. The introduction, and especially the methods and result section have
undergone revision.
2) The therapeutic interventions are mentioned pretty briefly under study population. This issue should be presented separately and expanded.

i. This has been published in detail in previous papers by Kongsted et al, however is now more clearly described in methods and now is shown in flowchart as well (fig 1)
3) The description of the treatments is confusing. Initially, three parallel treatment groups are mentioned; then, it seems that two treatments have been applied to the low risk group.
a. Please re-write and clarify. Has been done, see the above comments
i. Has been rewritten as above
b. I suggest that the treatments be introduced in the flowchart.
i. Has been done
c. The flowchart divides patients into two groups (low and high risk), but it is not clear what actually differentiates these two groups, since the right and left arm look almost identical. Please clarify.
i. Based on the risk criteria, the patients were divided into a high and low risk group. Being female however was decided to give 1 risk point based on other studies, therefore more women are in the high risk group (although being female was not a significant risk factor for 1 year work disability in neither the previous prospective study (Neurology 2001, Kasch et al) or this material (Eur J Neurology 2008, Kasch et al)
4) The palpation test that you used for the clinical assessment is of unclear origin.
i. ACR criteria by Wolfe et al, Arthritis Rheum, vol 33 160-172) has been applied as well as previous publication reference by HK added in the text
1. The results of this examination may be very examiner-dependent. We agree upon that.
Are there any data on its inter-rater reliability? We did training courses, in which all project/research nurses, physiotherapists, and doctors trained standardized palpation technique, pressure algometry and CROM measurement before the commencement of the study. Measurements (approx. 5 persons/examiner) were performed on thirty healthy subjects and with a short time-delay 1.5hrs -2 hours) reexamination was performed. During the timecourse of the project, the physiotherapists and project nurses met several times to calibrate the examination technique and discuss eventual problems. However, there were significant difference in variability in scoring patients regarding palpation in the centers but also pressure algometry/PDDT and PTT by means of 2 way Anova (Stata: anova “palpation_sumscore” center (A orB) Stratum(1-7) center#stratum.
For the main clinical measures: no significant differences for VAS score of max headache/neckpain ($p>0.20$) or number of non painful symptoms ($p>0.58$) and neither for cervical range of motion ($P>0.19$), and stratum (1-7) did not turn out as a significant factor together with center in any of the analyzes (e.g. center#stratum).
We have given some figures and made a short comment on the above in methods/statistics and results and discussion as well.
5) Primary outcome: please specify whether all the three criteria or only one have to be fulfilled in order to categorize the outcome as handicap.
a. Only one of three is necessary, the text is clearer now in “Outcome measures”.
6) The statistical analysis is not explained, which is unusual for such a complex study.
Statistics have been enhanced. We only apply initial measures and 1 outcome for 7 strata, therefore no longitudinal examination is needed
a. What kind of non-parametric statistics has been performed?
i. The risk factors were found in a previous study in which we examined by Kaplan Meier and Cox regression analysis and time dependency was also evaluated for the studied factors (Kasch et al Neurology 2001) similarly in 2008 Eur J Neurology this method was applied
ii. The statistical task is therefore more simplistic here, where we have an initial value for the measured parameters (pressure algometry, palpation, work related VAS scores and so forth and use a robust non-parametric Kruskal Wallis test grouping with use of the 7 risk strata.
iii. Data are (may be a little confusing, but has been explained in more detail) as well presented in graphs if normal - or log normal distributed as mean +/- sem, but for the statistics the K-W test was applied. K-W test also applied for the other analyses in which risk strata were analyzed. We only present data from 1 time point either the start or the ending (recovery +/-; number of days on sick leave).
iv. There is an inborn design weakness of the study, because we divided patients

into different treatment groups. However there was no substantial effect on outcome in neither low risk patients (verbal or booklet) nor high risk patients (verbal info; McKenzie phys, Semirigid neck collar), furthermore we did a subanalysis on the treatment groups in the high and low risk (original division in the project). And by means of stratification into their respective risk strata and splitting into low/high there was no difference in the treatment groups either (K-W p >0.15 (the high risk patients) and p>0.91 (the low risk patients), this has been added in results.
b. What methods have been used to analyze the influence of the different predictors on the outcome? ROC curves are now provided for each risk factor in suppl fig. 1
c. How did you analyze the effect of the three treatments and their potential confounding influence?
i. See above answer (7.0)
d. After I have read the results, it seems that you compared the different strata for the different parameters. This can certainly be done, but prognostic studies are typically analyzed by more complex procedures, e.g. multivariate analyses. Please comment.
i. The risk factors we have chosen were shown in previous studies (Kasch et al Neurology 2001 and Eur J Neurology 2008). Co-variance between non painful symptoms and painful symptoms are present and GLM analyses showing these calculations have previously been provided and we have previously shown an age, but not a gender effect on neck mobility. We therefore consider the situation different in this study applying this rather simplistic approach, but with a robust KW analysis.
7) At the end of the results, pressure algometry is reported, bus this procedure is not presented in the study aims and is not described in the methods section.
a. Has been specified in methods
8) It would be relevant to know more on the clinical performance of the stratification that you used. This is typically done by computing
a. sensitivity
b. specificity
c. and likelihood ratios of the tests.
i. Supplementary table 1 gives these data, as well as supplementary ROC curves for each parameter (CROM_negativized, Number of non-painful symptoms; VAS max-Headache/Neckpain and total Risk score)
9) At the end of the discussion you mention problems related to categorizing patients as being at high risk. The message is obscure to me, please clarify. In this respect, see my comment above: as long as the predictive value is not quantified, see suppl table 1 and ROC curves.
10) , it is hard to define the role of the stratification for clinical decision making. It can still be said that knowledge on the prognostic factors can help selecting patients for studies on preventive strategies, since it makes more sense to enroll patients who are more likely to develop chronic pain. I suggest that you stress this point.
MINOR ISSUES
I do not understand what segregated means, see e.g. the sentence in the abstract: Bio-psychosocial factors were significantly segregated from the first assessment by risk strata . Please check if this is a correct english term.
This term has been removed in the abstract and text as well
Under key messages, please delete the term more or less and specify what you mean by out of work .
Has been deleted (more or less) and (out of work) has been rewritten.
Under strengths , the third item would actually fit under limitations , since the need for further validating
the score is not a strength.
We agree, this third item has been placed under limitations.
Please remove more or less from the 1st sentence of the introduction.
Has been removed.
Exclusion criteria: I guess that the significant past pain conditions were exclusion criteria, which is not clearly specified.
The section has been totally rewritten for clarity.
Also, neck pain and headache is mentioned under the bullet list and again few lines below, this time specifying the VAS. Please correct and delete the repetition.
Please explain the abbreviation

CROM. Has been explained in Methods Clinical assessments.
The CROM scale indicates that the higher the score, the worse the points. Please explain the score for those
who are unfamiliar with it.
CROM details given in Methods Clinical assessments. And scoring system in detail under Risk Stratification in Methods.
Reviewer: Samuel McLean, MD, MPH
Vice Chair, Research, Department of Anesthesiology Attending Physician, Department of Emergency Medicine University of North Carolina, Medical School Wing C, Chapel Hill, NC, USA
This is an excellent and important study.
Reviewer: Dr Alison Rushton
Senior Lecturer Physiotherapy / Physiotherapy Research Lead University of Birmingham United Kingdom
I have no competing interests with this study.
There is valuable data here to inform assessment and management of patients following whiplash injury
that is important to publish. I would urge the authors to consider the feedback to improve clarity of the study and its clinical messages.
Thank you for the opportunity to review this interesting and large prospective study. There is valuable data
here to inform assessment and management of patients following whiplash injury that is important to publish. However the article does require considerable attention before it is acceptable for publication, as
clarity of methods and results are difficult for the reader to access in its present form. In particular, further
discrimination of content will assist clarity of the results and key clinical messages.
Abstract
1. The abstract contains abbreviations that affect clarity. It would be best to use all words in full for the abstract.
Abbreviations have been removed
2. The results need to be more clearly reported in relation to the defined outcomes of interest. The results need to accurately reflect the main text.
The result section of the abstract has been changed to reflect the main text
3. There is no analysis / discussion of findings within the abstract.
Has been added
4. The conclusion needs to relate specifically to the defined outcomes of interest.
Has been rewritten and shortened
The point re bio-psychosocial issues is unclear.
Has been changed in results of the abstract section
Article summary
5. This summary needs to reflect accurately the content and terminology of the finalised article.
The summary has been shortened and revised.
6. The detail of the RCT interventions in the limitations section is not relevant. It would be clearer to keep the focus to this study.
We have rewritten, and more clearly made reference to previous papers covering the treatments
Introduction
7. Is the premise that the Spitzer WAD classification was designed to predict outcome accurate? This is
also mentioned in the conclusion
The attempt from the original spine paper (The Quebec Task Force) was to extract best evidence what was known about relevant factors to describe whiplash patients from early after injury. The
time-scale from the taskforce group was not generally accepted/applied, however the WAD grading system
is applied in several prospective studies and also used in units in various countries to our knowledge, being
a gold standard.
8. What is the quality of the existing literature in this area? In 1995 when the Quebec task force did its

search on literature only few quality studies had been performed. In a round table discussion 2011(Sterling,
Carroll, Kasch, Kamper and Stemper, Spine vol 36, 25S, dec 2011, S330-S334) the prognostic factors of
whiplash injury are discussed, concluding " the current evidence is not sufficiently robust to be able to confidently predict outcome after whiplash injury" however a set of consistent risk factors are proposed
being priority measures for inclusion in future prognostic studies (Table 1). This reference and its message
has been mentioned in the Introduction
9. The previous study (Kasch et al, 2001) and the development of the tool to assess risk merits further consideration in this section to inform the reader.
References have been made to help the reader. Further elaboration on the subject added.
10. It is unclear why the other factors identified in the literature were not included in this study (first paragraph P11). This appears to be due to the timing of this study with data collection preceding this later
work.
This is a correct observation Impact of event was not studied in the previous study,
perceived injustice is a relatively new concept introduced by M Sullivan. We have from this study reported
on IES and emotional distress (Kongsted et al) from Symptom check list, SF-36 subscales and so forth (Tina
Carstensen et al, and other relatively new studies by Buitenhuis et al, McClean, Sterling et al have looked at
stress/distress, impact of event.
The more recent literature is therefore best included in the discussion for evaluating this study s findings.
We have according to above comments moved some of the introduction to
the discussion area.
Materials and methods
11. The design of this study within the RCT merits further consideration. Are there any implications of this design (multiple interventions across two trials) e.g. any potential treatment effects for your conclusions? Beyond a couple of brief mentions this point is not addressed.
In statistics and results this has now been considered, and statistics on this are provided.
12. This is not reporting a trial and therefore the Consort checklist is not appropriate.
We agree, the consort checklist has been removed
13. The clarity of this study as distinct to the trials needs to be clearer throughout. Much of the content re the trials can be removed as it is not relevant to this study. At present the inclusion of trial information is
confusing for the reader.
The section has been rewritten, shortened and hopefully more clearly described.
14. P12 refers to a group of low risk patients (line 31) and this is unclear. In looking back at the previous
trial it refers to allocation of low risk patients to this trial (Kongsted et al, 2008) following an allocation scoring system. The high risk patients were allocated to a different trial (Kongsted et al, 2007). Please clarify
and discriminate this content so that the reader is clear.
Has been rewritten.
For example can this be clarified within the inclusion criteria? Are you referring to participants of both previous trials being included in this study?
Yes, see changed flowchart for clarification and methods
If referring to participants from both trials is this risk categorisation relevant to the current
study? Has been clarified.
15. It is unclear why such a broad range of WAD grades were included (I-III) (P13). This requires justification. Were WAD classification 0 patients excluded?
Yes. They should present with relevant symptoms developed within 72 hrs after injury
16. The exclusion of significant headache or neck pain is unclear (P13). The rationale and detail of this
requires explanation. Does the later point (P14, line 22) link to this criterion? If so why were these patients

excluded? Has been explained in text in methods in new section “study population”
17. What is meant by “significant past pain conditions were in detail”: and the list afterwards? If (P13/14). If having a significant past pain condition patients were excluded.
18. The validity of the total risk assessment score merits mention and weighting of individual components is not mentioned (P15).
Are the included ROC curves from the previous study or current data? This requires explanation and if from
the previous study a reference rather than reproduction of the curves would suffice.
ROC curves have been provided from the present study as supplementary material Fig S1AD, as well as a supplementary table with likelihood ratios for each stratum. (suppl table 1)
19. There are differences in the description of outcome measures used throughout the paper which is confusing for the reader. The terminology and clarity of outcome measures needs to be clear throughout.
The outcomes of interest at 1 year set a priori are detailed as: handicap, NRS neck pain, NRS headache, and
Copenhagen disability (pages 18/19) but this is different to the abstract and is confused by the clinical assessment content and results sections where other outcomes are reported.
The term handicap has been removed as endpoint in the text. Abstract has been changed, Outcome measures have been rewritten.
Can the clinical assessment section be deleted (does not appear relevant to this study?)
We have made changes, but the assessment of CROM, palpation and pressure algometri eventually needs a
brief introduction for the general reader.
and the results section focused to the outcome measures of interest?
This has been rewritten.
The section has been rewritten according to these suggestions with reduction of “clinical assessment” and
introduction of “Baseline registrations, however for the graphs presented we consider the presentation of
palpation score,algometry crom measurement important for the reader.
20. The distinction of primary and secondary outcomes is relevant to the trial but not this observational
study of risk factors.
Secondary outcome measures are only briefly mentioned however the reader should know about the work
disability used in outcome.
The follow up dates of 3 and 6 months from the trial are not relevant to this study as it focuses on 1 year
follow up.
data was used in assessment of sick leave and recovery (measuring primary outcome measures, this has been briefly mentioned in methods
Results
21. The number of participants included needs to be clear on P21. We have removed information on the
subgroup with previous neckpain that we followed, see flowchart changes and removal/change in start of
result section.
22. The risk categorisation into low and high risk is again mentioned here (P21) and in the flow diagram
and appears irrelevant in this study see point 15. The results of the previous trial are not relevant to the
results of this study.
According to comments by professor Curatolo we have however put all detail in the flowchart, but reduced and simplified the text in the methods/results.
However, we have to inform the reader that this study was embedded in a randomized trial splitting into low risk and high risk patients receiving different treatments.
23. No detail is provided of the risk stratification in this section and the numbers in the strata.
Details are now present both in supplementary table 1
24. This section needs to relate specifically to the a priori defined outcomes e.g. why is handicap not reported as on P18? New outcomes are introduced in this section e.g.IES, assessment of physical job

demands, pressure algometry etc. Some outcomes are not reported e.g. NRS neck pain and Copenhagen
disability. This section needs to be considered further to avoid the potential of selective outcome reporting
/ data mining. In order not to give the impression of data mining also outcome for Copenhagen neck disability and neck/headache/shoulder and global pain, McGill pain data are reported in results. Initially we hesitated about bringing pain data forwards being part of the initial scoring system with eventual redundancy, but we agree upon your arguments.
Discussion and conclusions
25. There is no discussion of the findings in the context of the existing literature? This has been applied in
discussion where appropriate. The discussion on the risk factors and other potential risk factors has been
broadened
26. The point on P24 lines 37-45 is unclear. A lot of time is spent discussing the previous study on the risk
assessment score that would be best in the introduction.
The text has been edited in the discussion and some placed in introduction
27. The points re division of patients into two groups (P25) are unclear as the patients were not in two groups for the analysis and presentation of results. Is this point about potential treatment effects and best
included under limitations? Eventual grouping/treatment effect is discussed and statistics/data provided
28. Introduction of bio-psychosocial factors at this discussion stage is unclear as it is not mentioned earlier in the paper. The conclusion re bio-psychosocial is also unclear. This has been rewritten and psychosocial term has been removed.
29. Sample size for this study and number of participants in the strata are not mentioned. Added in supplementary table1
30. The exclusion of patients not working prior to the injury for the main outcome of interest needs to be
acknowledged as a key limitation. How many participants were therefore excluded from the analysis? Data
have been provided in EUR J Neurol 2008. H Kasch, they are relevant here as well, 30 were unemployed
(but were accounted for as “job available”, 10 with either disability pension or pension were excluded from
analysis. Has been added in results.
31. How are findings from the literature consistent with this study s results? (P27 lines 27-34) Is now discussed.
32. The clinical messages from the paper can be clearer to assist the reader.
Issues of presentation
33. The writing style, grammar and meaning are not always clear. E.g. suffering more or less (P9). More
or less has been removed, much has been totally rewritten.
34. Investigators should be blinded rather than blind-folded (P19). This slip has been removed

VERSION 2 – REVIEW

REVIEWER	Prof. Dr. Michele Curatolo Head, Division of Pain Therapy University Department of Anaesthesiology and Pain Therapy University of Bern, Inselspital 3010 Bern, Switzerland I have no competing interests.
REVIEW RETURNED	30-Nov-2012

- The reviewer completed the review but made no further comments.

REVIEWER	Rushton, Alison niversity of Birmingham, School of Health and Population Sciences, College of Medical and Dental Sciences
REVIEW RETURNED	16-Dec-2012

GENERAL COMMENTS	Thank you for resubmitting this article that was a pleasure to read. The valuable data is now presented in a clear and accessible format and the key messages and limitations of the study are clearly identified to inform the reader. A few minor points can usefully be addressed at the final stage of editing prior to publication: 1. The primary outcome is described differently throughout the abstract and text. Be consistent in the terminology to avoid confusion for the reader. 2. There is limited detail of methods in the abstract. Specifically risk stratification needs to be clear. 3. The 'baseline measurement' section (page 9 onwards) would usefully be termed 'risk stratification index measures' and detail the neck pain and headache, non-painful complaints, and CROM measure details to inform the score. The other measures could then be included under a heading of 'outcome measures' that will cover primary outcome measure and secondary outcome measures. 4. Page 14, lines 25-35 are unclear re the n=15 participants excluded due to protocol violation and n=52 participants being excluded after inclusion when they have a prior history of neck pain <4. 5. Page 14 needs to also include detail of the risk stratification score outcome. e.g. reporting how many in each group. 6. Page 18 mentions cold pressor and muscle strength that were not measured? 7. The flow diagram can be clearer: a. clarify that the baseline evaluation is for risk stratification. b. delete the intervention groups row as it relates to the two trials. c. detail assessment points as '3 months outcome assessment' etc rather than '3 mnths quest / interview' etc to be consistent to the text. d. Clarify the exclusions that are detailed after participants have been defined as eligible and consented. 8. Some minor issues of formatting and presentation that will be addressed at editing stage e.g. referencing.

VERSION 2 – AUTHOR RESPONSE

Thank you for resubmitting this article that was a pleasure to read. The valuable data is now presented in a clear and accessible format and the key messages and limitations of the study are clearly identified to inform the reader.

A few minor points can usefully be addressed at the final stage of editing prior to publication:

1. The primary outcome is described differently throughout the abstract and text. Be consistent in the terminology to avoid confusion for the reader. Has now been changed to 1-year work disability in all appearances in main text and abstract, besides more general discussions about prediction of work disability.

2. There is limited detail of methods in the abstract. Specifically risk stratification needs to be clear. The abstract now includes more information on risk stratification.

3. The 'baseline measurement' section (page 9 onwards) would usefully be termed 'risk stratification index measures' and detail the neck pain and headache, non-painful complaints,

and CROM measure details to inform the score. The other measures could then be included under a heading of 'outcome measures' that will cover primary outcome measure and secondary outcome measures. A very fine suggestion, and we have now changed the headings in the methods section according to this. All outcome measures have been reorganized in the text, starting with 1-year work disability and subsequently presenting psychometric and clinical measures. Inter-and intratester variability was moved to statistics, as it was more logically presented there after the revision

4. Page 14, lines 25-35 are unclear re the n=15 participants excluded due to protocol violation and n=52 participants being excluded after inclusion when they have a prior history of neck pain <4. In the main study we excluded patients with "pre-existing significant somatic or psychiatric disease, known active alcohol or drug abuse, and significant headache or neck pain (self-reported average pain during the preceding six months exceeding 2 on a 0-10 box scale, 0=no pain; 10=worst possible pain)." See pg 6 Study population. However, we chose to follow 52 pts (Average Neck VAS during last 6 months from 2 - <4) who
otherwise complied with the low risk group and they were treated according to the low risk group (RR for 1-yr work disability 4.6 !). 15 participants who were excluded due to protocol violation: wrong treatment group was given (information gave rise to different stratification into high and low risk group, typically pts with previous significant pain who underreported n=8, but also wrong initial stratification made by the study nurse n=7), has been explained in text under results, pg 14.

5. Page 14 needs to also include detail of the risk stratification score outcome. e.g. reporting how many in each group.

6. Page 18 mentions cold pressor and muscle strength that were not measured? Reference 19 Kasch et al SPINE 25S has been added, in which data from a previous study present data on cold pressor test and neck strength for extension and flexion. These figures are now supplied with this document, for your orientation (but not to be published here)

7. The flow diagram can be clearer:

a. clarify that the baseline evaluation is for risk stratification. "Risk Stratification at baseline" instead of "Baseline evaluation"

b. delete the intervention groups row as it relates to the two trials. The intervention groups row have now been deleted

c. detail assessment points as '3 months outcome assessment' etc rather than '3 mnths quest / interview' etc to be consistent to the text. The changes have been made as suggested.

d. Clarify the exclusions that are detailed after participants have been defined as eligible and consented. Details about drop outs, and exclusions have been added

8. Some minor issues of formatting and presentation that will be addressed at editing stage e.g.
referencing. We will of course comply with this.