Author's response to reviews

**Title:** Subgrouping patients on the basis of their individual course of low back pain over a six month period

**Authors:**

Iben Axén (iben.axen@ki.se)
Lennart Bodin (lennart.bodin@ki.se)
Gunnar Bergström (gunnar.bergstrom@ki.se)
Laszlo Halasz (laszlo.halasz@capio.se)
Fredrik Lange (fredrik.lange@bredband.net)
Peter W Lövgren (peterwlovgren@yahoo.se)
Annika Rosenbaum (annika.rosenbaum@telia.com)
Charlotte Leboeuf-Yde (clyde@health.sdu.dk)
Irene Jensen (irene.jensen@ki.se)

**Version:** 3 **Date:** 10 March 2011

**Author’s response to reviews:** see over
Response to Reviewer 1.

Major Compulsory Revisions

The statement on page 6 that the external validity of the sample has been validated is difficult to interpret, particularly since the reference for this statement is not in press. Does this statement indicate that the sample is representative of Swedish individuals with back pain? Or perhaps Swedish individuals with back pain in chiropractic clinics? Please clarify.

We realize that the statement on external validity was unclear. It has been clarified under “patients” in the methods-section. The general health score (EQ-5D) of our sample was compared to that of 1) a normative Swedish population as well as 2) that of a population awaiting back surgery. Then, the course of LBP over time was compared to the general course found in a systematic review (Pengel, Herbert et al. 2003).

Please provide additional detail on the measures collected at initial evaluation (page 7). In particular please indicate how self-reported sick leave was collected (number of days? yes/no? etc.) and how self-rated health was evaluated. The details of these measurement procedures cannot be accessed from the reference provided.

We realize that this has been poorly explained. We have added details of the initial patient evaluation under “data” in the methods section.

It appears these data would be best suited to growth modeling procedures used to identify latent classes, or a hierarchical mixed model approach. Can the authors provide further justification for the appropriateness of their data analysis strategy?

Again, this has clearly been poorly explained. We realize that the chosen method is not commonly used and may be new to this field of applications. The methods suggested by the Reviewer are excellent methods per se but the hierarchical mixed model does not apply to the research question here, that is, to find clusters of subjects with similar development of pain over 26 weeks. They are, with our vocabulary, variable-oriented methods (see our reference to Bergman and Magnusson, 1997). The latent class analysis has its merits and might be a suitable alternative. The method chosen by us also has its merits and in particular so for this data set and the aim of our study. The argument for choosing cluster analysis based on spline regression is added to the text under “methodology” in the discussion section.

The exclusion of participants who were not compliant with at least 80% of their text messages may be a substantial source of bias in the results considering that 27% of the sample was excluded on this basis. The analytic strategies suggest in the previous point are robust to missing values, and the ability to include a higher proportion of subjects as well as those with a constant response, may represent additional arguments for an alternative approach to the analysis.
The baseline characteristics of excluded subjects should be compared to included subjects.

A clarifying argument for only using highly compliant responders is added to the text under "the study sample" in the results section. With the extended explanation of both the cluster method and the use of regression lines to describe the course over time, the choice of using the high compliers only in our analysis will become more easily apparent. The previously mentioned analytic strategies will not give estimates of the individual subject’s course of pain so even if they are robust to missing values they will not be acceptable alternatives to the research question here.

As suggested, a comparison of patients included and excluded from the cluster analysis is now added, in the text under “the study sample” in the results section and in Table 1.

The potential confounding of the amount of treatment received is not reported on. How many subjects were continuing in treatment beyond the 4th visit? Did this differ among the clusters?

We do not have exact data concerning treatment, and can therefore not use this as a distinguishing variable between clusters. This is now clarified and discussed as a potential limitation in the discussion section under “methodology”.

The comparison of groups for the variable of total number of bothersomeness days does not seem to be appropriate considering that the number of bothersomeness days (the dependent variable) was also used to form the independent variable (ie, the clustered grouping).

Our poor explanation seems to have resulted in a misunderstanding. The number of bothersome days was not used as an independent variable to form the clusters, only curve parameters were used. We felt that the total number of bothersome days was an interesting variable to report for all clusters, as it would add to the differentiation of them. This is now explained more clearly in the text under “analysis” in the methods section. Further, in the now added multivariate discriminant analysis we restricted the included variables to be those strictly recorded at baseline. The reviewer’s argument is appreciated here.

Minor Essential Revisions

Page 6, I believe I understand the distinction, but the statement on page 6 that subjects were ineligible if they had chiropractic care in the past 6 months seems at odds with recruiting subjects currently receiving chiropractic care. This statement may be more clear if the authors indicated that the exclusionary criterion was no chiropractic care in the 3 months prior to the current episode (presuming my assumption is correct).
You have understood this correctly, patients were included at the 2nd visit of the current episode. However, this needed clarification in the text under “treatment” in the methods section. There were in fact two inclusion criteria. The first, to at all be considered for the study was to have LBP but not had chiropractic care within the past three months. The second criterion demanded that the patient returned for the second visit.

Page 7, please move the data on the time to the 4th visit (69% within 14 days) to the results section. In the methods just describe when this follow-up was designed to be obtained (on the 4th visit).

You are right. This sentence has been moved and the 4th visit is now instead mentioned in the methods section under "data".

Page 12, please move the basic descriptive information for subjects to the beginning of the results section.

As suggested, this section has been moved to the beginning of the result section.

The figures seemed to be mis-numbered. Figure 5 is referred to first in the text of the manuscript.

We understand your confusion. Figure 1 was referred to first, Figure 5 second. The reason was that we thought it would be easier to have Cluster 1 described in Figure 1, Cluster 2 in Figure 2 and so on. Figure 5 contains estimates of all the four clusters and is therefore mentioned after each of the other figures. However, in order not to breech the manuscript rules, Figure 5 is now referred to only after mention of all the individual cluster examples in the results section.
Response to Reviewer 2.

• In light of the fact that subjects are categorised on the basis of symptomatic course I have difficulty making sense of the decision to include all subjects regardless of duration of symptoms. Given that symptom severity drops sharply after initial assessment in 3 of the 4 clusters and the study design does not allow us to conclude that changes in symptom level are due to the intervention, how do we know where (temporally) a subject is along the course of their condition when they first attend treatment? I question whether an inception cohort should have been recruited for this study.

• The study lacks a coherent message with respect to how membership of the individual clusters can be predicted.

You are absolutely right; inference about treatment effect is not possible from this study, as is stated in the discussion. We have now addressed this further under “clinical significance” in the discussion section.

Concerning the predictive value of our clusters, we agree that cluster belonging cannot be predicted from the available baseline variables. However, this was not the purpose of the study. Our approach (using weekly data over a six month period) is merely a first step towards identifying subgroups. We have shown that clusters exist and it was also possible, with the modest extra information that we had, to find that these clusters had some clinical characteristics turning them into clinical subgroups. Obviously, our clusters will need further study with the addition of suspected baseline variables, both clinical and psychosocial characteristics, as is stated in the discussion section. Our work is a mere beginning, a starter so to speak, on further subgroup work. Therefore, it was not the purpose of this study to look for predictors of these clusters. However, the multivariate discriminant analysis now added to the analysis section strongly indicates that cluster membership might be predicted with a low error rate even from our small number of baseline data. What is needed to be developed is a more specific operational rule as stated above.

Major compulsory revisions

1. Please justify the decision to use only patients that supplied >80% of their data.

We realize that this was poorly described. A further argument for only using highly compliant responders is added to the text under “the study sample” in the results section. We hope that, in the extended explanation of both the cluster method and the use of regression lines to describe the course over time, the choice of using the high compliers only will become apparent.

2. The authors used linear regression lines to describe two stages of course for each cluster, did they ensure the data met the assumptions associated with this analysis?
Regression analysis was used to describe two stages in the course of pain for each patient. However, as no statistical testing of the regression parameters was done the usual normality assumption is not an issue. Linearity was checked by inspecting each subject's 26 weekly scores and how the subject's two regression lines could approximate the course. Evaluation of R-square was added. This is now clarified under “analysis” in the methods section to ensure that this misunderstanding will not arise again.

3. Please state how many subjects were recruited to the study.

The total number of patients recruited to the study is now added in the text both under “patients” in the methods section and under “the study sample” in the results section as well as in Table 1.

4. Please provide information about the baseline characteristics of the subjects included in the analysis and those who dropped out/were excluded for comparison.

The total number of patients included in the study is now added in the text both under “patients” in the methods section and under “the study sample” in the results section as well as in Table 1, as are their baseline variables, with a comparison of their baseline variables between included and excluded patients. We have also described the drop outs. In the discussion section, under “methodology”, this is further illuminated.

5. Given that approximately 1/3 of subjects were excluded from analyses perhaps a sensitivity analysis could be conducted to determine if inclusion of these data impacts the findings.

As the analysis here is conducted in several steps including spline regression, cluster analysis and discriminant analysis, a sensitivity analysis in the usual way it is conducted is probably in this study an impossible task or at least a very demanding task that probably in the end will add very little information. In a more conventional variable-oriented analysis we appreciate the value of a sensitivity analysis.

6. Pg 12-13, the two slopes describing the course of cluster 1 look to be very similar, did the authors conduct a test to determine whether they are actually different or likely different just by chance in this sample?

You are right, for Cluster 1, the “stable” cluster, the regression lines describing the early and later parts of the course are, indeed, similar. However, in Table 2, the difference between the two is visible. Our argument for not going further on the issue raised by the reviewer is that the figures are intended to describe features found in the clusters, not to test if a further simplification can be introduced in any of the clusters. For cluster 1, our “stable” cluster, an obvious choice is should then be to use a straight line with no slope.
7. Pg 13, paragraph 2; I am puzzled by the finding that such a high proportion (80%) of cluster 1 described themselves as definitely improved during the period that the mean scores on the main outcome measure are largely unchanged. This may suggest that the outcome chosen is not clinically meaningful to the patients. This finding needs to be discussed.

You are right. We have added text to the discussion under "clinical significance" of Cluster 1 concerning the clinical meaningfulness of the variable "number of days with bothersome pain".

8. Pg 13, paragraph 4; the finding that the cluster with the highest initial pain rating had the best prognosis is contrary to the vast bulk of prognostic research into painful musculoskeletal conditions. This raises the question of the meaningfulness of the selected outcome, alternately it may suggest a more serious problem with the analysis. The authors should discuss the reason for this diversion from previous research findings in some depth.

This is, indeed, an unusual finding. The rapid improvement of Cluster 2, despite the poor self-rated health and high bothersomeness, is discussed under "previous research" in the discussion section.

9. Pg 14, last paragraph; I miss a summary of how the clusters can be identified on the basis of baseline characteristics. This paragraph states only that some clusters are different from some others in age and initial pain intensity. It seems the study lacks a worthwhile message in terms of how membership to the clusters can be predicted.

You are right, these clusters cannot be predicted from the available baseline characteristics. This was not, however, the purpose of the article. We have clarified this further in the beginning of the discussion section. However, the multivariate discriminant analysis now added to the analysis section strongly indicates that cluster membership might be predicted with a low error rate even from our small number of baseline data. What is needed in the future, is a more specific operational rule as stated above.

10. Given that approximately 1/3 of the subjects originally recruited were excluded from the analysis, there is reason to be cautious about the reliability of the findings. Please provide discussion of this in a limitations section.

We realize that this was poorly described. A discussion about the exclusion of the poor compliers is added under "methodology" in the discussion section.
11. Please add the duration of pain to Table 1

Duration of pain the previous year is seen in Table 1. As stated in the methods section, no categorization according to the duration of their present LBP was used in the study. We have added further text to the discussion under “methodology” to enlighten this further.

Minor essential revisions

12. Pg 6, paragraph 1; Please reword or explain what is meant by the sentence, “The external validity of this sample...”.

We realize that this needed further elaboration, and the sentence on validity is re-written. It now reads: The external validity of this sample has been found to be acceptable, i.e. the general health and development of pain over time of the sample was compared to that of relevant populations.

13. Pg 15, paragraph 3; the authors state that cluster 4 had longer standing pain, this wasn’t reported in the Results.

Cluster 4 had a large proportion of patients who stated that they had had pain for more than 30 days the previous year. This was stated in the result section and now further clarified in the discussion under “clinical significance”.

14. Pg 15, paragraph 3; Please reword the final sentence of this paragraph for clarity.
15. Pg 16, paragraph 2; please reword the final sentence of this paragraph, it is unclear what the authors mean.

As suggested, these sentences have been reviewed. They now read: “This should make clinical management easier, as treatment may otherwise be considered disappointing to patients who are experiencing a stable course or slow recovery. Thus, knowledge about likely course is important for patient education.” and “Thus, compared to previous studies, this seems like an illogical finding. It is possible that highly bothersome LBP influenced the attitudes towards general health in this group of patients.”
Response to Reviewer 3

First, we would like to acknowledge the time and effort spent in reviewing our work. However, we want to make the reviewer aware of the fact that the checklist that he has applied for reviewing our article is more relevant in assessing a randomized study, whereas our study is a prospective observational study, not an RCT.

Yes the question is well defined, I am however not convinced by the relevance of it.

Regarding relevance of the topic, we think it is highly relevant to search for subgroups among patients with non-specific LBP. Todays’ situation, where patients are treated according to the theories of the provider they seek care with, has led to small treatment effects and high costs associated with “doctor shopping”. If subgroups could be identified, then treatments could possibly be tailored to individual needs. However, the search for subgroups is tedious and, so far, has not been outstandingly successful. This novel method of data collection has given us access to a new approach when looking for subgroups; detailed individual courses over time. From this development, we have categorized relevant subgroups. Any clinically relevant method to look for subgroups needs to be taken seriously.

Some have reported on the appropriateness of subgroup analyses. My issue with this study is, is that it is a description of a characteristic that can not be determined at start of the study. It is a description of “subtrajectories”, not so much of different characteristics of “subgroups”. In the end we learn that different people recover differently. The significance of the difference is clear, but the relevance of the difference is unclear.

Can a clinician predict the future trajectory of his/her patient based on characteristics present at first visit (the 4th week variable might be of interest if you want to know the trajectory at end of treatment only).

These outcomes become relevant if we are able to predict who should be classified to which trajectory and it would be interesting if this information could be used to improve treatment. The only use of such a classification is patient education and improvement of care. Only if therapy will differ substantially in the different trajectories and in the end the treatment effectiveness of the entire population increases will this be relevant. These issues are not addressed.

We see your point about prediction of cluster belonging, and agree that the clusters identified in this article cannot be predicted from the available baseline variables. This was not, however, the purpose of the article. This is merely a first step towards identifying subgroups, and our clusters will need further exploration to establish relevance in future hypothesis driven studies with the addition of suspected baseline variables such as clinical and psychosocial characteristics, as is stated in the discussion section. As also stated in the article, already the preliminary results in our study are relevant to a clinician treating patients with LBP. First, it is clear that most patients improve by the 4th/5th week. Thus, this is a “stop” sign in clinical reality, a time for reflection if improvement is not substantial. Further, some patients belonging to specific subgroup with a
slower trajectory, can be identified. However, we would like to point out that the multivariate discriminant analysis now added to the analysis section strongly indicates that cluster membership might be predicted with a low error rate even from our small number of baseline data. What is needed in the future, is a more specific operational rule as stated above.

Reporting on the longitudinal cohort part of this study should comply with MOOSE standards or similar.

We are surprised by this comment as the MOOSE standards were developed specifically for the reporting of Meta-analysis of observational studies. As such, many of the items in this checklist will not apply to our study (e.g. search strategies, databases searched, assessment of study quality, publication bias). On the other hand, some of the points in the checklist will be pertinent to the reporting of any study. We trust that our methodological approach will be acceptable after various “remedial” actions following the suggestions of the present three reviews.

I could not find Table one containing information on functional limitations, pain intensity, age, gender, duration of complaints (maybe even work status and job demands!) and radiating pain

We are sorry that you could not find Table 1. It was submitted with the original manuscript. It contains information on baseline characteristics but we do not have access to functional limitations and work demands in our data material.

Sun et al in 2010 have reported a number of criteria to assess the credibility of subgroup analyses

The article by Sun et al addresses an important issue: the analysis of outcomes in randomized trials where subgroups are identified on the basis of a patient or intervention characteristic. However, our study is an observational one, aimed at identifying subgroups based on clinical development, not an RCT with analysis of subgroups in relation to an intervention effect. We have therefore chosen to ignore this material as we did not find it helpful for our particular work.

Is the subgroup variable a characteristic measured at baseline or after randomization?*

There seems to be some misunderstandings here. First, this is not a randomized study. It is a non-randomized prospective observational outcome study without control groups. Second we have not used baseline variables on which to cluster, but we have studied the clinical development of individuals over time.
Is the effect suggested by comparisons within rather than between studies?

Again, there seems to be a basic misunderstanding here. This is not a study of treatment effect.

Was the direction of the subgroup effect specified a prior? Is the size of the subgroup effect large?

Again, this is not a study of treatment effect. There is no treatment to compare. We have not tested any effects between subgroups.

Is the interaction consistent across studies? Is there indirect evidence that supports the hypothesized interaction (biological rationale)?

This is irrelevant for our study. We have not tested or reported any interactions.

It is unclear what the flow of participants looked like (maybe in the earlier paper, but I would like to see it here) and I need that to know about generalizability. The 80% response rule seems questionable and it is not clear how many subjects are lost through this decision.

We realize that this was poorly described. These figures are now presented more clearly in the text both under “patients” in the methods section and under “the study sample” in the results section as well as in Table 1.

The authors are mainly concerned about others concluding on effectiveness of chiropractic care,

We are not referring to or drawing any conclusions about the effect of care.

The clinical relevance of the findings are examined by univariate comparisons that are of limited value.

As stated above we have performed a multivariate discriminant analysis (now added to the analysis section) which strongly indicates that cluster membership might be predicted with a low error rate even from our small number of baseline data. Together with the univariate analysis this adds to the examination of clinical relevance.

Limitations in the methods are not addressed. Recruitment rates are not reported and therefore also not discussed. Loss to follow up is not addressed. Biological rationale is not addressed
Limitations concerning the use of highly compliant responders, recruitment rates and loss to follow up are now addressed in the discussion under "methodology", and there is a mention of the dropouts in the results section under "the study sample". A rationale is given for choosing to subgroup on the basis of pain course in the background section.

*Chen et al in Spine*

We could not find the article referred to. When searching in the Spine Journal for this author, 45 articles are found, none of which seems to be relevant for cluster analysis or subgroups.