Author's response to reviews

**Title:** Depression as a 2-year outcome predictor of lumbar spinal stenosis surgery: a prospective study of preoperative, 3-month, 6-month and 12-month follow-up phases

**Authors:**

Sanna Sinikallio (sanna.sinikallio@kuh.fi)
Timo Aalto (timo.aalto@kyyhkyla.fi)
Olavi Airaksinen (olavi.airaksinen@kuh.fi)
Soili M Lehto (soili.lehto@kuh.fi)
Heikki Kröger (heikki.kroger@kuh.fi)
Heimo Viinamäki (heimo.viinamaki@kuh.fi)

**Version:** 3 **Date:** 28 May 2010

**Author's response to reviews:** see over
To the Editorial Office,

Referring to your e-mail of 21st April 2010, we have made some revisions to our manuscript. In the manuscript, the changes are highlighted in yellow. Please also see in detail our responses to Reviewers 1 and 3.

With these revisions, we resubmit our manuscript for your kind consideration and hope that our manuscript will now merit publication in BMC Musculoskeletal Disorders.

Unfortunately, we find it impossible to satisfy all the requests of Rev 1, as they suggest a lack of genuine attempt to evaluate this paper in a neutral although simultaneously rigorous manner; he even shows discontentment towards the revisions we have made by precisely following his earlier instructions. Moreover, he brings up new details needing revision that he has not brought forward in his earlier commentary.

In this respect, we kindly ask for Editorial assessment.

Respectfully, on behalf of all the authors

Sanna Sinikallio, PhD
Kuopio University Hospital
Department of Rehabilitation
Tarinan sairaala
71800 Siilinjärvi
Finland
e-mail: sanna.sinikallio@kuh.fi

Responses to Rev. 1 (Kevin Spratt)
Reviewer’s report:

As indicated in my specific comments, in my opinion:
- the manuscript needs a sharper focus regarding the purpose and hypothesizes.
- the methods need to be more specific regarding how the 30% improvement rule was implemented.
- the authors need to consider the value of the median split approach.
- important data is available that has not been analyzed or reported that would help the reader better understand the relationship between depressive state and surgical outcomes.
- additional discussion regarding the potential for surgical recover or the lack there off affecting or effecting depressive state should be addressed.
- the predictive power of the depressive state variable is not supported by the data or, more accurately, the results not presented is needed to support some of the discussion points.
- The discussion does not address or acknowledge that large ORs for depressive symptoms are reduced significantly after adjustment and that, for the most part the effects associated with VAS-based success are minimal and for ODI-based success are much more apparent when success is based on he median compared to the 30% improvement rule.
- Minor Essential Revisions
  In table 4 there is a problem with the 95% CI of the OR of OSW > median in model 3.

—We thank Rev 1. for his extensive engagement in our paper. However, we find it impossible to satisfy all the requirements made by Rev. 1, as he seems to be highly discontent of our revision and approach, no matter how extensively we revise the paper. Even more perplexing is the fact that Rev 1. comes up with totally new areas needing revision. This being the case, we feel we are on a “mission impossible” regarding Rev 1’s meticulous commentary.

However, we have made some revisions according to some the comments by Rev 1:

Specific comments:

Regarding a 30% improvement in relation to preoperative disability and pain must be clarified or expanded in the methods to make it clear that in both cases the 30% improvement is defined in the same way.

This piece of important information has now been added in the Methods. First, a “theoretical” 30% improvement was calculated as a 30% decrease from the patient’s baseline VAS score and Oswestry score. This calculation was done for each patient separately. Each
patient’s actual 2-year score on VAS and ODI was then compared to the “theoretical” outcome. In case the patient reached less than a 30% decrease, he was classified as having a “poor” outcome regarding pain and disability (The other class being “good”). For example, if a patient scored 45 on preoperative VAS, the theoretical 30% decrease would be $45 - 0.3 \times 45 = 31.5$. If the patient’s actual 2-year VAS was, say 50, he would be classified as belonging to the “poor outcome” group.

Introduction.
Defining how success is defined is out of place in the introduction and belongs in the Methods.

We thank Rev. 1 for this notion. At this point we refrain from removing the short sentence on outcome definitions from the Introduction. This is for reasons of reader convenience and clarity.

Statistical Analysis section: The text here is still too vague for me to understand the rationale for the various model of surgical outcome.

(...) The summary of the models is incomprehensible, at least in part, because the outcome point is not clearly specified and the hypotheses of interest are not sufficiently clear.

We did add extensive new analyses to the paper as recommended by Rev. 1 in previous comments. In addition, we clarified the rationale in the last revision of the paper. Three of the four Referees seemed content with our revision.

(...) Might it be of greater interest to evaluated surgical outcome at each follow-up point relative to baseline and then evaluate the ability of baseline information to predict speed of recovery and to explore pattern of recovery over time.

I have little trouble believing that the readers interested in spine surgery would be interested to know this. In contrast, as your paper now stands, I’m not sure what questions you are asking want what the results that you are reporting mean and how these results would influence the a clinician’s practice.

(...) We find these comments perplexing. Indeed there are various possible approaches to outcome assessment, and our paper aims at offering a possible approach. The conclusions we draw are as follows: treatment models that include the use of depression scales and appropriate treatment practices throughout the
preoperative and early postoperative period are encouraged. However, RCT studies are needed.

Results.

Your summary of dural sac area in the results is not mentioned in the introduction or methods and, therefore, the reader is left to their own devices to regarding why and how and why these values were obtained.

Rev 1 is right in pointing this out. For consistency, we omitted the dural sac result all together, as it is more thoroughly analyzed and reported by other researches of our LSS research group.

After telling me something that the intro and methods did not prepare me for, you then don't tell me anything about the outcome measures. Although the percentage of patients at or above the median is likely to be close to 50% for both the VAS and ODI (depending on distributional factors), what percentage of patients demonstrated a 30% improvement at 2 year follow-up on the VAS and ODI? Furthermore, why no table showing the bivariate relationships of the outcomes with the predictors? Without this information ORs can be misleading. Although Table 1 tells me the percentage of patients in a depressive state at each follow-up point, it does not tell me about the pattern of depression across time at the patient level. For example, how many patients were depressed at all assessment intervals? With 5 assessment times, there are 32 possible depression patterns, and I'm guessing fewer actual patterns. In summarizing martial status, it would appear that single, divorced and windowed patients were lumped together in the cause of making marital status a binary variable. Is this a good idea?

The univariate logistic regression models (regressing, separately, each of the 2 versions of the two outcomes on each predictor separately produces unadjusted Odds Ratios for each predictor. In my experience, with the outcome rate is low, ORs and RRs are reasonably similar. When the outcome rates are higher, which is the case in this study, ORs and RRs can be quite different, which, will affect interpretation.

We do acknowledge the point that there are several ways of interpreting the suitability of different statistical methods, and that this also applies to the selection between the use of ORs and RRs. However, in this study we chose to use logistic regression and ORs, which is a viable approach with this kind of data, also according to our statistician. At this point, the debate here appears to be heading to philosophical discussion about the preference of certain statistical methods over others. Despite the importance of such postulation, we would find it important to keep the focus on the actual topic of the paper, practical evaluation of the role of depression in LSS surgery, and apply methodology that is generally used and accepted in this area of research. As
we added several new tables and data in the previous revision of the paper, we feel that at this point, adding more tables and more statistical particulars is unlikely to bring any additional value, and could result not only in unnecessarily expanding the length of the paper, but also in a loss of focus.

As for the classification of marital status, we kindly refer to numerous studies on the importance of spousal support in protecting against depression (e.g. Kohn et al. 1998, Weissman et al. 1996). As far as we are aware, there are no data on the specific effects of different forms of living alone in the context LSS surgery outcomes. Moreover, due to the relatively small n of participants, a more fine-grained classification would result in loss of adequate statistical power and render group comparisons unreliable.

Discussion

(...)  

In logistic regression, the magnitude of an OR may not be a particularly powerful indicator of the quality of prediction. In linear regression predictive strength is about variance accounted for, In logistic regression, although SPSS does provide a pseudo variance accounted for index, predictive strength is about positive predictive value and model fit is about area under the curve (the interrelationship between sensitivity and specificity). Without supplementing your logistic regression summary results with this information it would be difficult to evaluate the veracity of your “main finding” that depressive symptoms are a strong predictor of surgical outcome.

We kindly refer to our previous answer on the rationale of not adding more data to the paper. Rev 1 requests several additions, of which we have made some, but adding all of this data would result in an extensively longer paper. In addition, three of the four Reviewers seemed content with our presentation of the results in the previous revision.

By presenting both the MCID (30% improvement) and the Median split approach for defining success you create a problem in that an examination of your results makes is clear that how you define success often makes a difference in the relationship between outcome and predictors. In my opinion you have a reasonable rationale for the 30% improvement rule and very little support for the notion of the median change rule.

We do agree with Rev. 1 in that how success is defined makes a difference in the relationship between outcome and predictors. However, this is a common challenge faced by any scientist. As for
spinal surgery outcome measures, there are no "gold standards" for the minimal important difference, and in reality there are often several grades of improvement that can be considered to carry clinical relevance (Mannion et al. 2009). In the present paper, we offer the reader two different approaches.

Responses to Rev. 3 (Jeffrey Katz)

Reviewer’s report

1. As per my comments on the first review, the Abstract needs to have some actual quantitative results.

Corrected as suggested.

2. As per my comments and the comments of another reviewer on first review, the OR per Stucki point is uninterpretable. Use of a binary predictor is ideal. Providing the reader with the scale range (eg 0-4) would be helpful so that the change per one unit can be put into context.

The scale range of the Stucki scales is provided in the Methods section. In logistic regression, the predictors may be binary or continuous and after consideration, we used Stucki scales as continuous predictors. Since Oswestry and VAS scales were used as outcome variables in logistic regression, the methodology required their categorization. The use of continuous symptom scales as covariates does, however, have advantages such as optimized utilization of data. Thus, we decided to use the Stucki scale as continuous in the analyses. Finally, as the Stucki scales were not highly significant as predictors, we did not use categorization.
References:

