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: 2 Date: 18 March 2010

Author's response to reviews: see over
Depressive symptoms during rehabilitation period predict poor outcome of lumbar spinal stenosis surgery: a 2-year perspective

Sanna Sinikallio, Timo Aalto, Olavi Airaksinen, Soili M Lehto, Heikki Kröger and Heimo Viinamäki

To the Editorial Office,

Referring to your e-mail of 18th February 2010, we have made several revisions to our manuscript. In the manuscript, the changes are highlighted in yellow. Please also see in detail our responses to Reviewers 1, 2, 3 and 4.

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

Respectfully, on behalf of all 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:
Abstract Evaluated after reading the full paper:
Abstract reports results in the methods and equates a Beck Depression Inventory (a depression screener) with depression when scores are at 10 or higher

We have now revised the use of term “depression”; in the whole of the revised manuscript we refer to “elevated depressive symptoms” when pointing to BDI scores ≥ 10.

3 2 1 “Not yet fully understood” suggests that we will come to fully understand, which is not likely. I would say that the factors that influence outcome remain controversial.

Corrected as suggested.

3 3 All This is a common problem in this area of research. We define depression based on a screener and then thoughtout the text define the patient a depressed if the screener (Beck in this case) score is above a given threshold. These screeners do not constitute evidence for depression, but rather suggest that further evaluation regarding depressive state should be done.
No statement of a hypothesis specified in Introduction

Corrected as requested.

5 1 All There is a difference in how one selects a sample and what the target population is. Hopefully sample selection is done in such a way to provide a representative sample of the target population. In this case the implicit inclusion criteria is that the patient was a viable surgical candidate. What defines these criteria? Are they related to the particular type of surgery?

We have now elaborated our wording on this, as can be seen in the revised Methods and Results. The inclusion was based on both
radiological and clinical evaluation. This was a clinical sample representing preoperative LSS patients in ordinary secondary level care. The following information has been added to the revised

Results: Preoperatively, the mean dural sac area of the most stenotic level was 69 mm$^2$ in patients having BDI scores ≤9 and 66 mm$^2$ in patients with BDI scores ≥10 (p = ns). Thus, the severity of stenosis and depressive symptoms had no association. We consider our sample to have been representative of LSS patients needing surgical treatment.

7 Study Sample. This is a result. In the methods you need to specify the analysis plane and estimate the sample size needed to achieve the desired power given the analysis method and estimated or desired effect size.

We thank Rev. 1 for pointing this out. However, as our study design was observational among clinical sample of patients, it included no experimental manipulation of the study group or control group. Therefore, no such preceding power analyses are plausible.

Don’t tell me the sample size in the methods, the methods presume no results are yet available. Don’t tell me about study attrition in the methods. The Results section should begin with issued regarding obtained sample size and patient attrition. The critical reader will want to know if the attrition biased the sample and this is often not adequately evaluated. An adequate evaluation of sample bias due to attrition involves evaluating: 1) changes in the demographic and baseline characteristics of the full sample (not including those lost to followup) with the patients who were lost to follow-up. Are these two samples different in age, gender, comorbidities, and baseling outcomes? If so, analysis of the patients who were followed should be adjusted to account for these differences with the covariates set at the levels observed for the full sample obtained at baseline.

In this study, the attrition was low, but the reader should, never-the-less, be provided with the nature of the attrition
As requested, we have now placed the information regarding the sample characteristics in the Results. As suggested, we did perform drop-out analysis regarding the main demographic (age, gender) and clinical (number of somatic comorbidities, BDI score, Oswestry score) preoperative variables. There were no statistically significant differences between the groups (the drop-outs, the dead and the study group), and this piece of information has also been added to the Results.

Logistic regression indicates that the outcomes of interest were categorized. A median split seems like a bad idea, as it defines failure as a 50% rate. A more reasonable approach would be to define outcome as a change from baseline, determine the amount of change necessary to define a successful outcome and to include the baseline outcome as a predictor. In this way, gain is adjusted for where you started. (...). Others have defined success as some percentage of change (e.g., Ostello et al). Although this definition was not clearly defined mathematically, the notion of a 30% gain seems only reasonable when considered as 30% of possible gain. If you define success in this way, then logistic regression approaches will work.

As suggested, we did perform new logistic regression analyses defining a “poor” outcome as a less than 30% decrease in pain and functional disability from the baseline values. These new analyses and corresponding results and discussion are now presented in the revised Methods, Results (Tables 2 and 3) and Discussion.

However poisson regression with a log link will provide relative risk estimataes associated with predictors rather than the odds ratio estimates that will be provided when using logistic regression. In addition, since covariates are involved, you might like to evaluate the obtained relative risks based on different covariate profiles.

We thank Rev. 1 for this suggestion, and consulted our statistician about the suggested methods. Both the odds ratio and the relative risk compare the relative likelihood of an event
occurring between two groups, and both measures have their advantages and disadvantages (for reference: Simon 2001). There is no definitive consensus on using OR or RR, and in practise, both are used. After careful consideration and based on the recommendation of our statistician, we decided not to re-analyse the whole of the data. Instead, we performed additional logistic regression analyses with new outcome criteria, as stated before. With these additions we hope that the results are now more understandable to the critical reader.

9 1 All As indicated above, this is where you need to summarize your sample gathering and attrition bias results. This process will inherently include the descriptive statistics of the retained sample.

Point taken; we kindly refer to the revised Methods.

10 Discussion, I have not evaluated the discussion as I can’t interpret the results as presented.

We have now revised both Results and Discussion and hope that they are now easier to understand

20 TABLE 2 This table does not provide me with the information needed to adequately interpret the results. First, as indicated above, in a predictive model sense, I would prefer to see Relative Risks rather than odds ratios. (…)

Please see our previous answer regarding the same issue. We chose to use logistical regression analyses, as we think that the choice between the OR and RR is more a matter of each researcher’s opinion and standpoint rather than a definite rule, when dealing with an observational study design. However, in the case of RCT, Poisson regression and RRs would be the first choice of analysis. The odds ratio is used extensively in clinical medical research settings, although relative risks may be much easier to interpret. Indeed, when events are quite common, the odds ratio and the relative risk will be very similar, provided that the odds ratio
is close to one (for reference: Davies et al. 1998). Moreover, OR gives an advantage of directly evaluating likelihood for both the occurrence of an event’s (i.e., the dependent variable) failure and success, since the presented OR is an inverse of the opposite outcome. The same advantage cannot be gained by using relative risks, which require more complex calculations, making them less practical for the reader to immediately evaluate the presented outcomes.

In almost all real clinical studies, interpreting ORs as though they were relative risks is unlikely to change any qualitative assessment of the study findings. Substantial discrepancies between the odds ratio and the relative risk are seen only when the initial risk is high (Davies et al. 1998). However, in the context of lumbar stenosis surgery, the risk for a poor surgery outcome has been observed to be somewhat low (Weinstein et al. 2008).

Given the logistic regression approach followed here:
First, I would like to see both the unadjusted and adjusted odds ratios.

Unadjusted odds ratios for the separate predictors are now presented in Table 2. These results were in line with the final results.

Second, I assume that the ORs for the continuous predictors reflect a 1 point change. This is often meaningless. (…)

We thank Rev. 1 for this point of view. However, we are inclined to think that for a clinical reader, this traditional presentation of ORs in logistic regression results is plausible.

Finally, a “large” Odds Ratio or RR, even when appropriate, does not necessarily provide evident of a strong predictor. Model sensitivity, specificity and postivitna and negative predictive values across the various models are important aspects for interpreting model accuracy and strength of prediction. The methods and results of this study do not address these issues, unless tucked into athe
discussion, which I didn’t read since I couldn’t interpret the results.

Indeed, we do agree with Rev. 1 that statistical parameters cannot be directly interpreted as definite reflections/predictions of the clinical reality. However, the predictive value of depressive symptoms in this study sample emerged logically, not with a drastic change in ORs, whether the outcome was defined either traditionally (30% improvement from the preoperative status) or in a more stringent manner in a previous version of the manuscript (belonging to the better scoring half of the study population on 2-year follow-up). Clearly, RCT studies are needed.

Finally, for the sake of limited informational value, we did not extend our reporting to statistical particulars (model sensitivity…) of the sixteen regression models more than we did in the revised manuscript. With this approach, we hope to satisfy not only the critical reader but also the clinically oriented one. As only one of the four referees asked for the details of the regression models, we would prefer not to report them to a more fine-grained degree. However, if the reviewers or the editor deem these particulars imperative, we are willing to add them in the next version of the manuscript.
Responses to Rev. 2 (Pradeep Suri)

Reviewer's report:

(...) However, it must be noted that there are now multiple publications examining the influence of depression on surgical outcomes, including many contributions by these authors. Given that the authors have previously shown depression to be associated with poor early outcomes of LSS surgery, and that depression in the preop and early recovery phase is associated with poor surgery outcomes at one year, it is unclear what these further analyses of this cohort add to our current knowledge. In paragraph one it is stated that ‘it is not known whether depression plays a different role as an outcome predictor in different phases of the recovery process.’, however, it could be said that the prior publications suggest rather strongly that depression is associated with poor outcomes at any stage in the process. The rationale for this study must be explained in a manner that is more convincing in the introduction section; if this cannot be better presented, the study rationale would appear weak.
We thank Rev. 2 for pointing this out. In the revised Introduction the study rationale has been brought forward more clearly.

Minor Essential Revisions
Title- The title of the manuscript is unclear. The current version at once seems to state that the outcome is something at 2-years, and then mentions preop, 3 month, 6 month, and 12 month timepoints without mentioning 24 months. Please reconsider the wording of this.

Point taken; the title has now been changed to a more informative caption.

Abstract:
Results- The words ‘being depressed’ in the phrase ‘Being depressed was a strong predictor of the 2-year surgery outcome.’ is unclear and does not make clear at what time points ‘being depressed’ was a predictor. Please clarify

This has now been clarified, as requested.

Conclusion- A period must be added after the word ‘outcome’ in line 2 of the conclusions section of the abstract

Considered and done.

Manuscriptp.
6, line 5- How was self-reported walking capacity reported and recorded?
Please clarify or reference

A question regarding self-reported walking capacity was included in the study questionnaire. The patients gave an estimate of their walking capacity in metres by answering the question (in Finnish): How long a distance can you walk continuously without pausing on even ground?
p. 6 line 9- Please explain why only ‘The latter VAS measurements were included in this study’ and which measurements are being referred to.

p. 6- lines 16-25. The description of the Finnish version of the Stucki scale can be improved. In the current version, it is difficult to ascertain what is being described is three parts of the same scale. The description of each scale or subscale should begin in the same way, ie ‘The symptom severity scale was _____, the physical disability scale was _____, etc’. this minor reformatting will make this easier for the reader.

p.7 line 1 should read ‘The scale concerned with post-operative satisfaction only was included in the follow-up questionnaires’

Furthermore, why was the whole scale not included? The rationale can be stated, if relevant

These details have now been clarified in the revised Methods.

Statistical analyses

7, line 16

This line would be more clear if restated as: ‘Regression analyses were performed using the data for the final 96 subjects <WHO REMAINED AT> or <WHO COMPLETED> two-year follow-up.’

p.7, line 15- what is the significance of the words ‘(method: enter)’?

These points have now been clarified in the revised Methods.

p.7. How was the determination made to use these predictor variables? Although is clear for some of these variables such as age and gender, it is unclear for others. For example, why were stucki subscales of symptom/disability included but the subscale of satisfaction was excluded? A few words on the selection of covariates for the model building would be helpful.

The rationale for predictor selection is based on two a priori defined points: 1) the well-known risk factors for depression were included, and 2) the clinically relevant factors were included as
predictors in the models. The satisfaction subscale of the Stucki scales was omitted because we have analysed it separately in our previous paper on satisfaction on 3-month follow-up (Sinikallio et al. 2007). However, to avoid confusion, in the revision we omitted the lines regarding the satisfaction subscale, since it is not analysed in this study. We also consider that the study of patient satisfaction falls into a somewhat different category than studying disability and pain.

p.8, final paragraph- The wording of this paragraph is unclear and be stated better One possibility would be as follows: 'The surgical outcome was defined AS A DICHOTOMOUS OUTCOME by using THE median VALUE FOR the 2-year Oswestry (median: 24) and VAS-scores (median: 0). A “poor” outcome refers to the subject HAVING A SCORE AT OR ABOVE THE MEDIAN VALUE FOR these measures.'

Results

Table 2.
It is unclear why for Model 1 all the ORs for all variables are presented, but this is not done for the other models. Given the fact that one of the major reasons for this study is to establish the relationship of post-operative factors to eventual 2-year outcome, there may be some utility to presenting all the data- including the nonsignificant data- in Table 2.

We thank Rev. 2 for this suggestion. However, after thorough consideration, we chose to report only the statistically significant and depression-related odds ratios in Tables 3 and 4, due to the limited informational value of presenting several columns of nonsignificant odds ratios and confidence intervals. Furthermore, as we added two tables (Tables 2 and 3) in the revision, we think that for reader convenience, only statistically significant results need to be presented.
The statement of Gender and Marital status under the column ‘Factor list’ is unclear. The ‘factor list’ names of ‘Gender (male)’ and ‘Marital status (single)’ will more easily allow the reader to understand which value the OR pertains to. Why are p-values only presented for Depressed in Model 1 and not for the other factors? This does not seem needed given that the ORs are presented, and if needed perhaps this can be superscripted.

As stated earlier, we chose to report only the depression-related and all the other statistically significant ORs in the tables (there were not so many of these). This has been now more clearly stated in the revised Results.

Discussion

P 10 Line 9- ‘According to logistic regression analysis, belonging to the depression group was the only preoperative variable that significantly associated with a poorer 2-year outcome.’ This statement appears not to be correct as per the data presented in Table 2, where it appears that baseline age was significantly associated with poor ODI at 2 years.

We have now re-checked our analyses on this; even though the OR in question is 1.04, it is not statistically significant when defining significance as p being less than 0.05.

P11, first sentence- Please clarify this line. Are the authors saying that the observed improvement could simply reflect natural history? If so, they should say this plainly.

This has been now clarified, as requested.

Discretionary Revisions

Manuscriptp3,

lines 4-5- The wording of this phrase while acceptable could be improved, one possible suggestion would be ‘The effectiveness of surgery for LSS has
been found reasonably good with severe cases of LSS, but success rates vary considerably’
P9 Line 1 The word ‘altogether’ may be unnecessary to start this sentence.

Again, these points have been clarified, as requested.

Responses to Rev. 3 (Jeffrey Katz)

Reviewer’s report:

Minor essential:
The abstract must include quantitative findings. At present the REsults comprise simply of two lines of text, all qualitative.

In the revised Abstract a more comprehensive presentation (several new regression analyses were done) of the results has now been included. However, due to spatial limitations were chose not to report the numerous statistically significant odds ratios with confidence intervals and probability values that we verbally summarize in the Results section of the Abstract.

Why did you choose to create a binary outcome? Why not take advantage of the full range of outcome scores on the Oswestry or VAS?

We chose to use logistic regression analyses to offer the reader some idea of clinically useful cut-off values for defining the outcome.

As in the Abstract, the REsults section of the manuscrip is too brief and not
sufficiently quantitative. Please report the actual OR's associated with depression in each model and indicate whether any other variables were significantly associated with the outcome.

We kindly refer to our previous answer to Mr Suri. We have now clarified the information regarding the presentation of the ORs in the revised Results and Tables: only the significant ORs and those related to depression are presented in the tables. Nevertheless, as also stated before, we are ready add this data if the editor and reviewers find it necessary.

Major compulsory:

These data do not support a recommendation to treat depression pre or post operatively. The data do suggest that research investigating treatment of depression in patients undergoing surgery for stenosis merits a high priority. You should address this.

This is an essential point to consider. We have now corrected our conclusions accordingly in the revised Discussion.

Table 1: The model building and display could be more informative. You should examine the bivariate association between depression and outcome and than the multivariate OR, adjusted for other covariates. This will provide insight into potential collinearity among different concepts (eg functional status and depression).

Explain in the table that the regression coefficient for the Stucki scores can be interpreted as Oswestry points per Stucki point.

We kindly refer to our answers to Mr Spratt; we have now added new data to Tables 2 and 3.

Clearly, the Oswestry score will be predicted by pain and functional status, especially when the latter are measured postoperatively. The question raised by this research is whether there is additional explanatory value provided by
depression. Can you comment on this? The value of using depression in screening would be if it provided independent data not already offered.

Again, this is an important point to consider. With the added new analyses and results we think that the additional explanatory value provided by depression becomes even more pronounced. It is well-known that depression and disability are closely intertwined, but they are indeed also separate phenomena. We think that the accumulating data on the biological underpinnings of depression also pertain to the depressive symptoms of LSS patients. We added some new considerations on this to the Introduction and Discussion.
Responses to Rev. 4 (Gustavo Zanoli)

Reviewer’s report:

Overall comments:
The authors should be congratulated: this is an interesting study that falls within the scope of a musculoskeletal journal. It is a prosecution of a long-lasting and fruitful study so not entirely new, but the methods are well-established and sound. Little work is needed before it can be considered, in my opinion, useful and acceptable for publication, even though the degree of interest for the readers should be assessed by the editorial team.

We thank Rev. 4 for these words.

1. Is the question posed by the authors new and well defined?
The question is well defined but not new, however authors justify the reason to publish this study. Maybe this can be made more explicit for readers

Considered and done, as can be seen in the revised Introduction. 
The prospective assessment in relation to the outcome is our new idea.

(...) 

6. Do the title and abstract accurately convey what has been found?
I think they can be made more informative especially differentiating what is new in this paper (even knowing the other ones, it took me some time to grasp).

We have now revised the Title and Abstract to make them more informative, as suggested.
Methods

1) VAS: It is a common rule that patients should fill in outcome instruments alone, without external interference. I wonder why of all VAS measurements, the most biased one was chosen.

We are sorry for our previous confusing wording on this in the Methods, which has now been clarified: Overall back and leg pain intensity was assessed with the visual analogue scale (VAS: range 0-100 mm) when visiting the study physician as a part of the study protocol, without any interference.

2) Cut-off for depression: I am not familiar with the BDI instrument, it seems to me that if 10/63 is abnormal, it might have a significant ceiling (or floor) effect. Do the reference you quote justify your choice? I believe you should expand on this point for the more general readers.

We have now expanded the rationale for this cut-off in the Methods. The BDI was originally developed to measure the depth of depressive symptomatology; hence higher BDI scores indicate more depression-related symptoms (cognitive, motivational, somatic, emotional). The BDI has been found highly reliable, valid and discriminative. The cut-off 9/10 for detecting depressive symptoms was originally suggested by Beck and Beamesderfer (1974) to be used among medical patients.

3) Study sample: I think it would be useful to know how sample size was calculated, if it was related to the present aim of the study or if it was taking into account other expected outputs of the study, published elsewhere or not.

We kindly refer to our previous answer to Mr Spratt: Preceding sample size calculations are not relevant in observational study designs.
4) Surgical outcome. I would have rather used absolute measures of success rather than relative ones. If I understand correctly, median split means that there is always 50% belonging to the poorer scoring half. This might not reflect clinical outcome, as usually LSS patients do well in more than 50% of the cases, at least in the first 2 years. Why not simply using absolute score, of improvement above a predefined MCID? It would be interesting to know how result would look like, this way.

Again, we refer to our previous answer: As suggested, we did perform new logistic regression analyses defining a “poor” outcome as a less than 30% decrease in pain and functional disability from the baseline values. These new analyses and corresponding results and discussion are now presented in the revised Methods, Results (including new Tables 2 and 3) and Discussion. Essentially, the results remained the same.

5) Predictors: Did you explore other possible predictors? For instance, surgical severity or peri-intervention adverse events, or surgical achievement (fusion, decompression) as judged by the surgeons themselves? If not, please comment, or at least address in discussion.

These other important predictive factors have indeed been studied and reported by other researchers in our study project. Our main objective was only to examine depressive symptoms and related factors.

However, we did perform some preliminary analyses regarding surgical approaches and depression: In the surgery, 19 patients were treated with spondylodesis (sp) with or without instrumentation. Preoperatively, the mean BDI score was 8.8 among patients treated with sp and 10.6 among patients without sp. On 2-year follow-up, the respective BDI mean scores were 7.2 (patients with sp) and 7.7 (patients without sp). As the differences in the mean BDI scores between the two groups gained no significance preoperatively or on follow-up (p = ns. both preoperatively and on
follow-up), we deemed unnecessary any further analysis on this matter in this paper.

Results
6) Expand: This section is really succinct, maybe it could be usefully expanded, for example adding more data or exploring alternatives such those suggested in point 1) and 3).

Considered and done.

Discussion
7) Personally, I am not completely convinced by your interpretation of the data, at least the ones you present here. How can you say what comes first, symptoms or depression? It is a common experience that chronic patients tend to be depressed, but many of them would improve after solving the organic somatic problem, as it happens in your sample 48% depressed preop vs 29% postop with no specific treatment for depression. It would be interesting to be able to differentiate among these 2 types of “depression”. Does BDI do that?

As we have now hopefully clarified in the whole of the revised manuscript, the possible causal order of symptoms/depression/ - depression/symptoms may be irrelevant, as there is accumulating evidence that the biological mechanisms of depression such as immunological disturbances are associated with wound healing and pain among surgical patients (Starkweather et al. 2006, Kudoh et al. 2001, Cole-King et al. 2001). Therefore, these two “types of depression” may not exist at all, as this biological bidirectional mechanism between depressive symptoms and body functioning may the essential issue.

8) Did you consider alternative explanations, such as that those patients which were correctly operated did better and THEREFORE are not depressed anymore and the others which did not improve in their symptoms, also did not improve
depression. In this case depression could simply accompany symptoms, being present pre- and postoperatively in a subset of persons who are prone to depression when they experience clinical problems...please make your point more clear if you want refute this idea.

We kindly refer to our previous answer. We have now expanded our discussion on this in the revised manuscript.

9) I strongly disagree with the idea that your data “pinpoints the need for ...appropriate treatment”, especially if you mean medications. I believe this statement is not supported by your findings and should be addressed with a specific RCT, in case.

Rev. 4 is correct in pointing this out. We have now attenuated our wording on the clinical implications of the results. Indeed, RCTs are needed.

10) As you point out, it is curious that only a minority of patients received medication compared to the high % of patients considered “depressed” with the BDI cut-off? Again, I am no expert (I am a surgeon), but why do you think this is necessarily a mistake? Couldn’t they receive other types of treatment? How does this compare with normal prescribing behaviour in Finland? Is there a commonly accepted guideline?

As our study was a prospective observational design among a clinical sample of surgically treated LSS patients, psychiatric assessment or treatment was not included in the study protocol. However, the study patients’ use of medication during the study was enquired. When looking at the reported frequencies of antidepressant use, one would be safe to assume that depression among this patient population may well have remained largely undetected. In addition, psychotherapeutic treatment is not likely to be the first choice of treatment among back pain patients.
11) Page 11 line 16-17: “these result support the use of antidepressants”. I disagree. See pint 9. And I believe is outside the scope of this paper, I would simply delete it.

Considered and done.

12) Apart from suggestion in favour of medicaments, I also see another danger in your paper: surgeons will use depression instruments to exclude patients form the operation, in order not to ruin their “statistics”, Did you consider that? Do you think it would be ethical to do so? I believe not. If depression is a clinical problem people should receive appropriate attention, but if they also have a real physical problem, they should get the right intervention without discrimination!

We do indeed agree with Rev. 4 on this! We have now also briefly commented on this at the end of the Discussion.

------------------------------------------------------------------------------------------------------------------------
Minor Essential Revisions
Methods
1) WAI for co-morbidity. Modifications of validated instruments are to be avoided. Specifically, self reported number of illnesses could significantly underestimate co-morbidities in my opinion. Could you motivate your choice?

The WAI is a well-validated questionnaire for assessing working ability that is widely used in Finland. Instead of modifying it, we simply used one item of the questionnaire to record the self-reported number of illnesses. We do agree with Rev. 4 that subjective estimates of health and sickness may be prone to biases. However, in clinical reality, the patients’ subjective estimates need to be taken seriously if we are to satisfy the needs of both the patients and the medical professionals. However, as the number of co-morbidities was not analysed in this paper, we omitted it from the Methods.
2) Stucki: Your study, with all the data you have, would be a very good occasion to publish the Finnish validation with little additional effort. In my personal opinion this probably more urgent and useful than some of the information in your present study. I am sure you are planning to publish it, I just wanted to encourage you.

We thank Rev. 4 for this excellent suggestion.

Discussion
3) Page 11 1st line: 10-15% is the percentage of patients who get better, or is it their individual improvement?

It is the decrease in depressive symptoms: the symptoms decreased by 10-15% in the short term without treatment. We have now clarified this in the Discussion.

----------------------------------------------------------------------------------------------------------------------

references:


