Reviewer's report

**Title:** Predictors of Shoulder Pain and Disability Index (SPADI) and work status after 1 year in patients with subacromial shoulder pain.

**Version:** 3  **Date:** 21 July 2010

**Reviewer:** Danielle van der Windt

**Reviewer's report:**

The authors have taken many of the reviewers’ comments on board, and considerable changes have been made in the manuscript.

**Major Revisions**

1) General: The structure of the text and grammar could be improved, and there are quite a few typo’s, e.g.
   - complain = complaint (introduction)
   - sickness leave = sick leave (introduction)
   - physically work related factors = physical (discussion)
   - heterogeneity = heterogeneity (discussion).

   These are just a few examples. It might be good to ask a native speaker to copy-edit the manuscript, this may help to improve clarity and grammar.

2) Discussion: the sentence which forms the second paragraph (‘We included a clinically selected group with subacromial pain in a clinical trial which might contribute to the prognosis’) is a bit of a mystery. What does this mean? What does contribute to the prognosis? (selection of a subgroup of subacromial pain, or the fact that this is a trial population)? And does contribute to prognosis in a good or a bad way? Why?

**Results:** It might be good to reorganise the results section (and include subheadings)

- Study population (Table 1)
- Predictors of pain and disability at 12 months (including Table 2 – univariable analysis + Table 4- multivariable model for SPADI), and interaction of baseline SPADI with treatment
- Predictors of work absence at 12 months (including Table 3 - univariable analysis + Table 5- multivariable model for work absence)

**Minor revisions**

Methods: The authors explain why they used forward selection for building the logistic regression model: “only 10% of the lowest category of the outcome variable was allowed simultaneously in the model”. This sentence is still not clear.
to me. Does this mean that the analysis required 10 events (i.e. people not in work at 12 months) for each predictor entered in the model? If so, please make this clear.

This also means that no more than 3 variables should be entered in the multivariable logistic model. This must have been tricky for the model using work absence as the outcome measure. I would certainly not recommend to also include interaction terms in a model with this low event rate.

Results: Some of the results may need more explanation to facilitate interpretation, for example

• The odds ratio for general health was 1.06. Explain that this means that the odds of work absence increased by 6% for every point increase on the EQ-VAS (if that is correct...)

• Beta-values for the subgroups are now both presented (19 and 6.8), but these estimates should include 95% CI intervals (rather than p-values). I am also not sure if the interpretation of the beta-values (in the discussion section) is completely right: My thought was that this should be: … ‘For patients with high baseline SPADI scores the adjusted difference in SPADI scores at 12 months between patients receiving exercise versus shortwave was 19 points in favour of exercise, whereas this difference was 6.8 points in patients with low baseline SPADI scores.’

• ‘The logistic model was statistically significant...’ I am not sure at all about the meaning and relevance of this statement, and would omit this from the text.

Discussion: Based on the significant associations of perceived health status & education with work absence, the authors state that ‘this may confirm that cultural and psychological variables are important predictors of outcome’. I still don’t understand this conclusion... Poor general health is indeed a plausible predictor of future work absence, but how do cultural and psychological variables come into this? These factors have not really been addressed in the analyses, and there could be multiple alternative explanations for the association between health and sick leave... I would suggest taking out this statement.

Discussion: The authors have added the following paragraph: ‘Studies have attempted to identify the best predictors for pain relief and for return to work in patients with musculoskeletal pain. Our results may support the conclusion of van der Windt et al. [38] that distress seems to be more strongly associated with low back pain than shoulder pain.’

Thanks for referring to this study, but I am not sure if this is relevant here... The authors have not compared shoulder pain to back pain, and other studies in shoulder pain patients did find significant associations between distress and pain/disability in shoulder pain patients, so only referring to van der Windt et al. may not be correct. And why emphasise this one non-significant finding in the discussion section? There are others that may be more surprising, for example, baseline duration of pain was not significantly associated with outcome.

I would like to suggest taking out this paragraph, and not discuss individual
associations between specific predictors and outcome. In a multivariable prediction model, the emphasis is on prediction of outcome. If the authors are interested in associations between specific prognostic factors (e.g. psychological factors, education, or work load) and outcome, they should have formulated a different research question, and used a different analytical approach (building explanatory rather than predictive models).

Discussion: The authors indicate that the AUC (0.78) implies satisfactory discrimination between patients who were "working" and those who were "not working" after 1 year. The AUC is difficult to interpret and prognostic models nearly always end up with an AUC between 0.72 and 0.80. The risk of overoptimism is high, especially in prognostic studies based on a small sample size (there were only 23 events in this study). It would be better to re-organise this paragraph, as follows:

“Numerous predictors were investigated, but the final linear model explained no more than 30% of SPADI’s variance [17]. Similar results were found for the logistic model, which suggest that between 22% and 33% of the variability was explained by the two predictors. Heterogeneity of the patient group despite uniform clinical criteria; measurement error of the outcome variables; and prognostic factors not examined may contribute to the moderate percentage of variance accounted for [16,17]. The AUC of the model of 0.78 may be interpreted as satisfactory discrimination between patients who were working and those who were not working after 1 year. It should be noted however that a small sample size with a relatively high number of predictors investigated tends to overfit the predictive model and may spuriously overestimate associations between factors and outcome [45].”

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Not suitable for publication unless extensively edited

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests