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: 2 Date: 31 May 2010

Reviewer: Danielle van der Windt

Reviewer's report:

This paper describes the design of a prognostic model in a trial population of patients with subacromial shoulder pain. The study focuses on the combination of predictors that is most strongly associated with pain & disability at 12 months, and with work status at 12 months. This study has used good methods to develop the multivariable models, but the paper also has some shortcomings, mainly in terms of presentation and interpretation of results. It is important in these prognostic models to decide between either a predictive or explanatory approach, and this decision is not always clear in this manuscript.

Major Revisions

1) Introduction: The authors explain that several prognostic models have been developed for shoulder pain, but that most of these have concerned more heterogenous populations of shoulder pain. What were the expectations of the authors: had they expected different factors to be of importance in patients with subacromial shoulder pain? Is a prediction model within this subgroup the best way to investigate factors that are specific to different types of shoulder pain? wouldn’t that require a large heterogeneous population that represents different diagnostic subgroups so that interactions between predictors and diagnostic category can be investigated? This would possibly require a more explanatory analysis rather than a predictive approach as used in this paper. This issue could perhaps be addressed in the discussion section.

2) Main findings: The results show that a few general factors are the strongest predictors of outcome: education, previous pain, and current disability levels for pain & disability, and general health & education for work status. These are the usual suspects… So what is new about their study, and what does it adds to existing evidence on the prognosis of shoulder pain?

3) Interpretation of findings: The authors acknowledge that predictors found in their study are not specific to subacromial shoulder pain, and not even specific to shoulder pain. This deserves more discussion. What is it that nearly all of the many prognostic models in musculoskeletal pain produce the same set of predictors?

4) Sample size: The study size was fairly small for building a prognostic model. There are no strict guidelines for sample size calculations for such analyses, but when using an outcome measured on a continuous scale (pain & disability), a
rule of the thumb is to have at least 10 cases per variable in the model. The sample size (n=94) seems to be just about enough for the analysis of pain and disability, with about 10 variables introduced in the multivariable model. For the analysis of work status (dichotomous outcome), a sample size of about 10 ‘events’ for each variable in the model is usually recommended. But it is unclear what the event rate was in this study: what proportion of people was not working at 12 months follow-up? I expect that this proportion may be fairly low (20-25%?), which means that this model is likely to lack power and be more susceptible to the effects of sampling variation. I am not sure if the use of forward selection is the best way to deal with this problem.

5) Performance of the logistic model: The authors present the % correctly predicted, but this not very informative. If the prior probability of being out of work at 12 months is about 20% (with the proportion of people in work being 80%), a correctly classified proportion of 82% actually means that the model did not add much information at all, as this proportion includes the proportion of people correctly classified by chance. It is much better to present the AUC (C-statistic), and perhaps also the prior probability of the outcome (% not in work at 12 months) plus the distribution of posterior probabilities based on the prediction model.

6) Sampling variation: The authors rightly acknowledge that the small sample size and the use of a relatively large number of predictors may have led to overfitting and to overoptimism of estimates of the association between predictors and outcome. But this does not only mean that the models are “unsuitable for making inferences about the effectiveness of interventions and cannot imply causality”. Prediction models are by definition not suitable for these purposes... It also means that the results are likely to be affected by sampling variation, and – when tested in another population – will probably have much lower predictive performance. When presenting a prediction model based on a small sample size, it is recommended to include some sort of validation (either using bootstrapping techniques, but preferably by validation in another dataset), or – if that is not feasible – to be very cautious in terms of interpretation of the results. The authors need to discuss the influence of sampling variation – given their small sample size, and perhaps also the fact that this is a trial population – more carefully.

7) Interactions: Were all interactions between treatment group and prognostic factors tested? Were other interaction (between predictors) tested as well? This should be more clearly explained in the methods section (statistical analysis), including the rationale for testing these interactions. Why were these of importance in this predictive model?

8) Interaction: A significant interaction was found between baseline SPADI score and treatment group. The authors performed a stratified analysis for patients with low versus high SPADI scores to present separate estimates for the effects of treatment. This is possible, of course, but as this paper focuses on a prognostic analysis (and not on exploring subgroup effects of treatment in a randomised trial), it might make more sense to present the analysis stratified for treatment,
and describe the prognostic value of baseline SPADI score depending on the
type of treatment provided.

Whatever way the interaction is presented, the results should be given for both subgroups (not just one), including a clear explanation of the meaning of these results so that the readers understand the interpretation of a regression coefficient of 19.0.

9) Discussion: This study is based on a trial population. The authors realise that this may have drawbacks in terms of generalisability due to strict eligibility criteria and more selective response. They explain that refusal rate was low, and that a range of occupational groups were included which “suggest that their findings are also valid for patients in primary care”. This issue deserves more discussion. Eligibility criteria may also have affected generalisability, but more importantly: what about the interventions offered in the trial: supervised exercise or shock wave therapy? How may interventions influence the results of a prognostic study, and how may these interventions affect the generalisability of the study?

10) Values of individual predictors: When building a prediction model, it is all about the combination of factors that best predicts outcome, and predictive performance of the model as a whole is the most important outcome. (Confounders are of no importance, all predictors are equally important to start with - although it is indeed recommended to include treatment as a confounder when using a trial population). The association between a specific predictor and outcome does not have to represent a causal association at all; some variables may only be of importance in combination with other predictors. So caution is needed when discussing the individual variables in a prediction model. The authors may have over-interpreted some of their findings. For example, on page 8: “We found that a high score on EQ-VAS evaluating current health status predicted work status of ‘working’, which confirms that cultural and psychological factors are important predictors of outcome”. But of course, a lower score on the EQ-5D does not necessarily imply that cultural or psychological factors are of importance (low EQ-5D scores may be driven by poor physical function). Furthermore, when discussing the association between general health and work status, issues of comorbidity, coping, the role of physicians, and work conditions are all being discussed. This is interesting in theory, but the combination of factors that is the result of a predictive analysis may be greatly affected by characteristics of the sample (especially when sample size is small). This means that it may be better to focus the discussion on issues related to predictive performance of the model as a whole, sampling variation and generalisability, than on a detailed interpretation of the specific value of individual predictors. An analysis of the predictive value of individual predictors (adjusted for confounding by other variables) would have required a very different analytical approach and different presentation of results.

Minor Essential Revisions

1) Abstract: Make clear that the outcome measure of the prediction model for pain and disability concerns absolute values of the combined SPADI score for pain and disability. It is now not clear whether absolute values or change in pain
& disability is used as an outcome, and whether combined or separate scores are used for pain and disability (the latter is more common when using the SPADI).

2) Why did the authors decide to use the combined SPADI score, rather than investigate predictors of pain and disability separately? Previous research has shown that predictors may vary across different types of outcomes.

3) The authors explain why they used forward selection for building the logistic regression model: “because only 10% of the lowest category of the outcome variable was allowed simultaneously in the model”. Does this refer to the general rule that the analysis requires 10 events (i.e. people not in work at 12 months) for each predictor entered in the model?

**Level of interest:** An article of limited interest

**Quality of written English:** Needs some language corrections before being published

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

**Declaration of competing interests:**

I declare I have no competing interests