Reviewer's report

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

Version: 2 Date: 22 March 2010

Reviewer: 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

- Discretionary Revisions

No Comments

Specific comments:

Abstract: 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 (i.e., as the percent of possible gain), which is important because this implementation of this type of improvement depends on how the score is scaled. For example, If pain is scaled from 0 to 100 with 0 = worst pain state and 100 = least pain state, (like the BP scale in the SF36) then a 30% improvement from a score of 60 would be .30 x (100-60) = 12 points. On the other hand, if disability is scaled from 0 to 100 with 0 = no disability and 100 = worst disability (like the Oswestry Disability Index), then a 30% improvement from a score of 60 would be .30 x 60 = 18 points.

Introduction.

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

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. After further reading it now seems clear that the surgical outcome is based on 2-year change from baseline for Pain (VAS) and disability (ODI). For both the VAS and ODI high scores represent worse outcomes, and therefore the Improvement must have been based on Baseline – 2-year follow-up as a 90 at baseline – 30 at follow-up would represent a 60 point improvement. From the 30% change rule, the possible improvement from baseline would have been 90 points, and, therefore, for that patient a change of 27 points (.3 * 90) would have been required to classify that patient as having demonstrated a minimally important clinical difference. Thus this example patient would have been considered improved, as would, for example, a patient whose baseline score was 30 and whose 2 year follow-up score was 20 (mcid = .3 * 30 = 9).

My point here is that an example in the methods of how 30 percent change was computed and more clearly pointing out that the surgical outcome was based on change from 2–year follow-up relative to baseline would be useful to the reader. Especially since the Ostello et al paper does not provide such examples.

Having said this, your paper does not make it sufficiently clear if patient outcome is model independent. In other words, is outcome always based on 2-year follow-up relative to baseline or does the “baseline” shift to be consistent with the timing of the predictors (i.e., for 3-month predictors, the “baseline” that 2-year VAS and ODI scores are compared to are the scores obtained at the 3-month follow-up? This is not clear to me.

By providing 4 models – presumably predicting 2-year surgical follow-up (1/0: good/bad, improved/not improved) – each based on patient characteristics (age, gender marital status, Stucki scores, and depressive symptoms), where the
differences in the models are the Stucki scores and depressive symptoms since these are measured at each follow-up time interval, isn’t it important to test for differences between model results? Your statistical analysis does not suggest how you will do this nor does your introduction provide a rationale for what value there is in making these comparisons.

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. For example, maybe patients with depressive affect at baseline do not demonstrate as good a recovery at 3- or 6-months follow-up but show no significant differences at 1- and 2-year follow-up compared to patients without depressive affect, after controlling for age, gender and marital status.

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.

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.

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.
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.

A somewhat “picky” point here is also that the analyses you conducted were not “multivariate regression analyses” (which mean that you were looking a multiple outcomes simultaneously), but rather multiple logic regression, which means that you were looking at one outcome at a time with multiple predictors). This is a mistake common in the medical literature, but a mistake none-the-less.

You indicate that at baseline higher age was predictive of higher ODI scores. The OR reported in the table was 1.04 with a confidence interval (1.00 – 1.09). A significant p value is hardly powerful evidence for strength of association here. Although this OR is likely to reflect the increase in odds of being in the “bad” group if one year older, and if you computed the OR for 5 years change in age rather than 1 year age

Change, the OR would be 1.20, even then, reporting this result hints at over-interpretation of your results, which is never a good thing.

Although post operative depressive symptoms may indicate those patients at greater risk of a poorer surgical outcome, it also seems likely – or at least intuitive – that patients who do not experience recovery after surgery may be more likely to experience a more depressive state. This begs the question regarding how ODI and VAS changes at follow-up intervals predict reported depressive state. How many people who reported improvement from baseline at 3 months follow-up also reported lower depressive state>

These are simple correlations of the change scores at these intervals and would be important for helping to understand this “Catch-22” or “Chicken and Egg” problem.

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. The pros for the 30% rule (assuming you applied it as I did above in my example) is that the 0/1 status of each patients is directly related to the individual patient’s baseline and outcome score with different MCID values depending on the start point. In contrast the median rule is
based on the distribution of the difference scores and does not account for baseline. Thus, virtually all patients who did had reasonably low baseline scores (less ODI, less Pain) will be in the no success group, which means that the median approach is biased toward patients with larger baseline "bad" outcomes. When the outcome is consistent between the two rules, this looks good, when the results are not consistent, the critical reader should be doubtful regarding the potential value of the statistically significant outcome. When results don’t “line up” across outcomes, the researcher and clinicians should be sensitive to the possibility that the observed “significant” result may be a fluke.

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

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

- Have you in the past five years received reimbursements, fees, funding, or salary from an organization that may in any way gain or lose financially from the publication of this paper, either now or in the future? NO
- Do you hold any stocks or shares in an organization that may in any way gain or lose financially from the publication of this paper, either now or in the future? NO
- Do you hold or are you currently applying for any patents relating to the content of the manuscript? Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript? NO
- Do you have any other financial competing interests? NO
- Do you have any non-financial competing interests in relation to this paper?

I don’t know if they are competing interests but I am currently working on a manuscript that also is looking at the relationship between depressive state and outcomes in this same patient cohort treated operatively or non-operatively with many of the same predictors. I don’t view the publication of this paper as having any affect of the likelihood of the manuscript we have in preparation being publishable, and therefore, I don’t feel my review was influenced by my involvement with a similar manuscript. If anything, I would like to see this manuscript published since, as the authors point out, there is not an abundance of information in this field.

If you can answer no to the entire above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.
I declare that I have no competing interests