Reviewer’s report

Title: Searching for success...: Development of a combined patient-reported-outcome (“PRO”) criterion for operationalizing success in multi-modal pain therapy

Version: 4  Date: 19 January 2015

Reviewer: Christopher Harle

Reviewer’s report:

Major Compulsory Revisions

This paper addresses an important clinical problem, how to measure success in multi-modal pain therapy. As the authors motivate well, multi-modal pain therapy is widely viewed as essential in complex pain conditions, and insufficient research has gone into best practices for measuring success of multi-modal therapy. Important to the validity and usability of such a criterion, the authors focus the components of their success criteria on patient-reported and widely available instruments. However, as I detail in my comments below, I think the authors could greatly improve the manuscript with better conceptual justification for their success criteria and a stronger empirical validation.

1. It is useful that the authors generated two versions of success criteria. Moreover, the input variables seem to have reasonable face and content validity. However, leading into their empirical evaluation the authors should better justify, based on prior literature and measurement theory, the potential appropriateness of each version. For example, there are many ways to possibly aggregate the individual components of the success criteria to arrive at a dichotomous measure of therapeutic success. However, the authors simply present two such possibilities with little justification for their appropriateness relative to other possible aggregation schemes.

2. Building on comment 1 above, the authors provide a reasonable but not rigorous comparison of the pros and cons of Versions I and II. Moreover, this comparison appears in the discussion section rather than the results section. If a significant contribution of the paper is meant to be a comparison of two potentially-valid approaches, I recommend conducting a more conceptually and/or empirically driven comparison and presenting it as part of the results. Alternatively, if this comparison is not meant to be a key part of the contribution, the authors might simply describe how they arrived at their single preferred method (see comment 1 above) and then focus only on validating one of the versions in the remainder of the paper.

3. In addition to providing better conceptual justification for the success criteria evaluated, the authors should provide a more complete empirical evaluation of the version(s) they ultimately choose to examine. In particular, it would be
appropriate to conduct additional sensitivity analyses around the criteria for identifying yes or no success. For example, in Version II, the authors could adjust both the number of individual variables for which improvement is needed (i.e. 3 out of 5) and the number of standard deviations improvement (i.e., 1). Similarly, in Version I, they could consider looking at 2-point improvements. The choice of such sensitivity should be bounded by justification of what might be considered clinical relevant improvement.

4. Building on comment 3 above, as part of evaluating different operationalizations of yes/no success, the authors should consider including other empirical to demonstrating the construct validity of their success criteria. In the results section for research question 1, the authors compare their two versions and show they seem to correlate reasonably well. But, they do not provide any empirical validation relative to other measures to which the new measures should theoretically relate. For example, the authors could greatly strengthen their claims of a useful measure if they were able to conduct additional empirical validation, such as discriminant, convergent (e.g., does success correlate with a patient self-reported global satisfaction measure?) or predictive (e.g., does success predict future patient outcomes of importance?)

5. The authors mention that 42 patient observations were excluded due to 50+% missing data. In the remaining included observations, were there any missing data? If so, please explain how missing data was handled in your analysis.

6. Version II is described as a more conservative approach to measuring success, yet more people were judged as successfully treated than under Version I. Please clarify what you mean by “conservative” and how that coincides with the empirical results.

7. The results section should be re-written in a more streamlined and consistent manner. The section would be more readable if there were less description of process and interpretation with a focus on the quantitative study results:

Statistical results should be consistently presented with the relevant statistic (e.g., odds ratios) and p-values in the narrative, not just the tables.

The results section should avoid stating or re-stating methodological details. For example, essentially the entire first paragraph of the results, which begins “Two combined success criteria,” contains content more appropriate for the methods section.

8. The manuscript should contain a section describing the study limitations. In the current version, there is no limitations section.

9. While I think the segment of the paper that models the relationship between the authors’ new success criteria and various clinical variables is interesting, it’s not particularly clear how it ties in with the primary aims of the paper. Are the authors claiming that because their measure of success correlates with some clinical variables, it is therefore more valid? If so, please justify more.
Alternately, are the authors simply interested in identifying factors related to success? With this knowledge, for example, clinicians might then be able to predict, a priori, which patients are more likely to experience success? I think this is important, but it seems like a separate aim rather than part of the validation of the success measure.

Minor Essential Revisions

1. The authors’ writing overuses parenthetical descriptions, such as to remind the reader of a detail introduced previously. I encourage the authors to cut back on this practice and generally be more succinct. For me, this style actually made various sections of the paper harder to follow due to a lack of concision.

Discretionary Revisions

1. The authors utilize multiple sensitivity analyses and a reasonable process for selecting variables to include in their final model. However, they might also consider an automated approach to variable selection (such as a stepwise approach).

**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:**

I have received research funding from Pfizer, Inc. to study the use of patient-reported outcomes by clinicians treating pain.