Reviewer's report

Title: How well do clinical prediction rules perform in identifying serious infections in acutely ill children across an international network of urgent-access datasets?

Version: 4  Date: 10 September 2012

Reviewer: Peter Wyer

Reviewer's report:

General

The authors have made substantial revisions in the attempt to address this and the other reviewers’ comments. The resulting manuscript is more transparent and consistent and somewhat clearer in its objectives. The reviewer is not entirely persuaded that the limitations of several aspects of the authors’ approach, including the retrospective design and the approaches to missing data and to varying definitions of predictor variables across the data sets have been adequately acknowledged. Specifically, their assertion that varying predictor definitions are acceptable because ‘there is always variability in clinicians’ interpretations and assessments’ quite misses the point of structured, validated prediction tools in the first place. The whole point of emphasizing prospective validation of such tools is that clinicians MAY WELL differ in their interpretation and application of one or more predictors even when consistently defined and worded. Hence, for this very reason, performance of a prediction rule assessed via a retrospectively collected data set can only constitute a hypothesis with respect to performance when actually used by clinicians. This is also why prediction rule developers frequently go to great pains to select predictors that show a high inter-observer agreement in the derivation phase. Retrospectively generated results based on ‘approximations’ of variables across data sets must be considered to be entirely preliminary and very likely inaccurate or wrong. The authors have performed sensitivity analyses to try to address some of the limitations of their study as initially presented. However, they have largely limited these analyses to statistical comparisons of AUCs generated via ROC curves. AUC is a measure of accuracy of a rule or test and may not be predictive of clinical performance, particularly where stakes associated with false negatives and false positives may be high. Furthermore, it is unlikely that this study has been adequately powered to detect differences even within this parameter. The reviewer believes that this manuscript may be salvageable provided that the severe limitations of the methodology are clearly acknowledged, preferably in a separate limitations section preceding the discussion section. Additionally, important additional improvements in organization and clarify of presentation are called for.

Major issues still to be dealt with:

1. The manuscript needs a separate limitations section in which the many
potential vulnerabilities ensuing from retrospective comparison of potential rule performance across independent data sets are clearly acknowledged, with no attempt to down play them.

2. The criteria used to define “low”, “intermediate” and “high” prevalence settings need to be clearly identified in the methods section. This review overlooked the absence of this the first time around and apologizes for inefficiency.

3. The organization of the manuscript needs substantial improvement. The Methods (p. 6), for example, starts out with 2 paragraphs that largely reproduce content already presented in the introductory section. The sensitivity analyses presented in supplementary files need to be described fully in the Methods section. All primary quantitative results of the study, including those of the sensitivity analyses need to be presented under Results, not in the Discussion section. As noted above, a distinct Limitations section needs to be added and should frankly acknowledged the limitations of the methodology without attempting to dodge their implications. The Discussion needs to relate your results to prior literature and to broader context of application. Much of the discussion as written is inconsistent. For example, in one place you state that few prediction rules have been properly validated and even then are frequently not used. Later you assert that clinicians, particularly in primary care settings, will need and use prediction rules to guide their practice. Which is it? Finally, there is inconsistency in many of your recommendations. Is further research needed to identify new “predictors” of serious infection, or to prospectively validate the most promising prediction rules. If the latter, which specific instruments would you propose?

Minor specific issues

P. 5: The focus of the article could be substantially improved if you made clear here that most existing prediction instruments were derived for use in emergency care settings and that your intent was to explore potential applicability of such instruments to primary and ambulatory care settings. This section should end with a concise statement of your OBJECTIVE in performing this analysis. There is no such statement anywhere in your manuscript, including the abstract! (The reviewer apologizes for not having picked this up in the earlier version.)

P. 6, Par. 2: As per previous comments, the wording here would much more appropriately be revised to “prospectively assessed”, since the YOS was widely considered to have failed clinical validation for the purposes for which it was originally developed. I.e. it was not actually "validated".

P. 8: A discussion of your approach to sensitivity analyses should be inserted here. Even if those analyses were post hoc, this should be acknowledged and they should be described here.

P. 9, Par. 2: This implies that you considered pooling of data but rejected doing so due to heterogeneity. Please describe how you assessed heterogeneity.

P.13, Par. 2: As noted previously, AUC is a very insensitive way of assessing
clinical performance of a prediction rule or test. Please consider presenting your sensitivity analyses in the same format as your primary analyses, i.e. sensitivity, specificity and LR.

P. 14, Par. 2: The wording here is misleading. Your results almost certainly do NOT reflect clinical practice and are at least 2 to 3 level removed from it due to the issues raised previously. This is just one reason why a Limitations section PRECEDING your Discussion section might add appropriate perspective to your manuscript. The reviewer does not believe that an analysis of this sort can legitimately support clinical recommendations, only research recommendations.

P. 15, Par. 2: You present new quantitative data here and elsewhere in the discussion section that belongs under Results.

P. 16, P. 3: As noted previously, this line of argument is entirely specious. It flies in the face of all published methodological guidelines for developing prediction rules which uniformly stress well defined predictors and vetting through assessment inter-observer reliability (e.g. Laupacis et al JAMA 1997;277:488-494).

P. 16, Par. 3: While objectively true, the opening statement of this paragraph has to do with an issue far beyond the scope of your inquiry. It also contradicts assertions made later in this section to the effect that clinicians need and will rely on prediction instruments in this area.

P. 17, Par.2: The treatment of the YOS throughout the manuscript is disjointed and inconsistent. Are you proposing that, despite the fact that it was thoroughly considered for application to young infants but rejected for clinical use almost 30 years ago that your results suggest that it might be useful for different purposes (ruling in infection) and/or in different (older) populations? On a larger scale, the reviewer fails to detect a consistent and compelling line of argument through the discussion section. Perhaps debriding it of content that belongs in earlier sections will pave the way to greater clarity.

P. 17, Par. 3: You attempt broad clinical recommendations here that are in no way supported by your methods or results. Suggesting that practitioners should notice a child’s breathing, temperature and whether they look sick or not is a recommendation that does not require the kind of research you are summarizing and analyzing to arrive at. These principles are taught, and hopefully learned, long before a student moves beyond undergraduate studies! The last sentence of this paragraph makes an assertion that is extraneous to your study in a different way: Your manuscript in NO WAY addresses the performance of individual predictors in predicting serious infection or more specific infections. There is ample research that does exactly this. This is not what we have here.

P. 18, Par 2: The reviewer believes that you are on the right track in focusing on recommendations for research rather than for clinical practice, given your methods and results. However, your recommendations here are not supported by your findings. There persists a confusion (see previous comment) regarding
whether the focus of your inquiry is performance of individual predictors or of structured prediction instruments. Your data only addresses the later. You earlier in the discussion flirted with the recognition that most developmental efforts in prediction rule development stop short at the derivation stage, rendering clinical application premature. This reality is nowhere better illustrated than in the area at hand. A compelling approach to interpretation of your data could therefore start with this observation (which was supported by your earlier published SR), and proceed to make 'best evidence' recommendations regarding WHICH of the non-prospectively validated rules appear to be most potentially fruitful to pursue with prospective validation efforts, and in what populations.

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:

I declare that I have no competing interests.