Reviewer’s report

Title: A prognostic tool to identify adolescents at high risk of becoming daily smokers

Version: 2 Date: 5 March 2011

Reviewer: Chris G G Richardson

Reviewer’s report:

Overall, I feel that the research question is interesting, novel and of interest to readers of this journal. The methods are for the most part appropriate and well described though I have pointed out several specific issues involving the methods and an added cautionary note related to the limitations. With the exception of the aforementioned issues, I believe the manuscript is an interesting and well written contribution to the literature on the prevention of adolescent smoking.

Although I am not an expert in statistical modeling of this type I do have several methodological concerns that I feel could be easily addressed by the authors that would improve the work:

Major Compulsory Revisions

1. Approaches to dealing with missing data in Data analysis section:

The authors indicate that “if, within any given wave, data on daily smoking were missing for one cycle only (out of five), we assumed that the participant did not initiate daily smoking during that cycle.”

No reason is provided for why the authors take the aforementioned approach to imputing their missing data. Given the major advancements in methods of dealing with missing data and the accessibility of such approaches I suggest that the authors redo the analysis using a more rigorous technique (eg multiple imputation). Extensive literature exists on this topic – see for example:


Similarly, the authors indicate that “data on the covariates were drawn from the “baseline” cycle within each wave. We used the “first observation carried forward” and “last observation carried backwards” strategies to impute data for covariates missing at “baseline;” without any justification – despite being widely used and intuitive this approach is known to be problematic.

See for example


2. Testing for interaction effects

The authors make no mention of testing for interaction effects in developing their model – given that including interactions in the model would not increase the burden to users of the tool these should be at least tested as part of the model evaluation process.

Minor Essential Revisions

3. Provide a measure of the shrinkage

I appreciate the authors’ use of bootstrapping to correct for over fitting - it would be useful if they directly reported the shrinkage factor as an assessment of the bootstrapping approach.

4. Limitations on generalizability of the predicted probabilities

I think it would be useful if the authors discussed the potential impact of heterogeneity within their limitations. There is growing research supporting heterogeneity in smoking trajectories as well as the determinants associated with escalation in use – this may be an important area of research to consider when developing/validating prognostic tools. Validation of the model on a new dataset should also be done before widespread use of the tool is encouraged though I suspect that the probabilities would not change substantially – despite this I feel that a “use with caution” message be made very clear to readers.

Discretionary Revisions

One other point the authors may wish to consider discussing is the inclusion of intentions to smoke (i.e. assessed on some form of Likert scale) in future versions of the tool – assessing smoking intentions makes intuitive sense and the literature shows them to be strong predictors of subsequent smoking behaviour.

**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 declare that I have no competing interests