Reviewer’s report

Title: Evaluation of the Propensity score methods for estimating marginal odds ratios in case of small sample size

Version: 1 Date: 16 February 2012

Reviewer: Duminda Wijeysundera

Reviewer’s report:

Pirracchio et al. present a simulation study (with accompanying case analysis of a real dataset) to evaluate the relative performance of propensity score (PS) matching and PS inverse probability of treatment weighting (IPTW) with regard to Type 1 error, bias, variance, and mean square error (MSE) for estimating marginal odds ratios in smaller datasets.

In datasets ranging from 1000 to 40 observations, the authors assumed an event rate of 50% and assessed both a 50% and 20% prevalence of exposure. They varied several other factors, including:

• Strength of association between exposure and outcome (ORs of 1, 1.5 and 2.5)
• Choice of variables included in the PS model

The authors generally found that IPTW performed better than PS matching when sample sizes exceeded 40 subjects. Nonetheless, overall Type I error rates were less than 10% in all sample sizes that were assessed.

In addition, the authors found that

• Exclusion of true confounders from the PS resulted in increased bias – with the impact generally being greater for the IPTW approach
• Inclusion of factors only related to the outcome resulted in decreased variance
• Inclusion of factors only related to the exposure resulted in increased variance

Overall, this is a well-written manuscript describing a methodologically sound study. I have a few comments.

Major Revisions

While the authors assessed several key aspects of using PS matching and IPTW, there are 2 additional components that I believe merit assessment.

• The authors consistently used an outcome event rate of 50% in their samples. However, for most binary outcomes, such an event rate is rare (even for small sample sizes). I would suggest that they at least assess these same simulations when event rates are as low as 10% to 20% (which is often much more realistic for clinical studies).
• While the authors assessed the correct specification of the PS model, they did
not address the specification of the outcome model. For PS matching, this is a
not a major issue, since there a model-free approaches for calculation marginal
However, for IPTW, there is a concern that the outcome model relating the
exposure, covariates, and PS weighting is correctly specified. Indeed, IPTW
appears more sensitive to bias even when the PS model (as opposed to the
outcome model) is incorrectly specified. I would suggest that that the authors
also consider cases where a continuous covariate has a true non-linear
relationship with the outcome (e.g. exponential) but where the outcome model
specified only assumed a linear relationship. Alternatives could include
non-inclusion of true interactions. One would assume that PS matching would be
less sensitive to this problem provided that covariate balance is achieved in the
matched pairs. I am less certain how IPTW would perform. This is a practical
concern since it is difficult to be certain as to “true” relationship of all prognostic
variables with the outcomes in any specific clinical dataset (hence omission of
non-linear components or interaction terms is plausible concern).

Minor Revisions

In the “data-generating process” section of the Methods section, the word
“guarantee” has been misspelled as “guaranty”.

Level of interest: An article of outstanding merit and interest in its field

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