Reviewer's report

**Title:** On the use of propensity scores in case of rare exposure

**Version:** 2  
**Date:** 27 July 2015

**Reviewer:** Romain Pirracchio

**Reviewer's report:**

Review on 'On the use of propensity scores in case of rare exposure' by David Hajage, Florence Tubach, Philippe Gabriel Steg, Deepak L Bhatt and Yann De Rycke in BMC Medical Research Methodology

I read with great interest the manuscript from Hajage et al. The study addresses an important issue that has been pointed out as a potential caveat in many simulation studies on propensity score estimators.

Overall, the manuscript is very well written and has substantial merit. However, I do have some concerns with the paper in its present form.

**GENERAL COMMENT**

The authors chose to evaluate the performance of PS estimators of the ATT and the ATE in case of rare exposure. I am not really sure it is sound to estimate the ATE in this context. Indeed, this parameter has 1) to make clinical sense; 2) to be identifiable from the data. Is it realistic to estimate what would be the difference if the entire population was exposed versus non exposed, when, in fact, less then 10% of the population is actually exposed ? I think the only quantity of interest in this context is the ATT. In addition, there might be strong violation to the positivity assumption when addressing the ATE in this context. This might partially explain the fact that ATT estimators consistently performed better that ATE estimators.

**MAJOR COMMENTS**

- Page 4, line 6-8: could the authors elaborate on how the intercept delta 0 was actually selected?
- Page 4, line 36: the authors chose to focus on low treatment prevalence, but used quite large sample sizes in their simulations. The study would certainly benefit from additional scenarios with smaller sample sizes (at least down to 500), as PS estimators are frequently used in such settings;
- Page 5, line 38: given their simulation plan and the way the estimate the PS, the authors placed themselves in a very conservative situation, where the PS model is correctly specified. My guess is that any PS model misspecification could further impair the properties of PS estimators in the context of low exposure prevalence. A scenario with some minimal misspecification (such as some correlation between B and C) would be strengthen the conclusion of the paper;
o Page 5, line 22-24: I would suggest using The Abadie-Imbens standard error estimator. To the best of my knowledge, it is the only existing SE estimator that has correct coverage when the true PS is used for matching, as it takes into the uncertainty related to the matching procedure by itself. Other ‘robust’ SE estimators based on the influence function do not take into account the matching procedure and thus has suboptimal coverage in this context;

o Page 6, section “performance criteria”: my major concern is that the other did not select the mean squared error among their performance metrics. In the context of PS estimators (especially with IPTW), there are many situations where the bias might be limited but the variability of the estimator might blow up. Therefore, the MSE, which accounts for both bias and variability, should therefore be included in this set of metrics;

o Page 6, line 25: the authors state that they performed 5,000 sets of simulations, but is it the overall number of simulation or the number of simulation where there was enough event in the exposed group?

o Page 6, line 37-40: could the authors elaborate on how the true ATE and ATT where computed?

o Table 2: I am not sure to understand n stands for? The sample size should not be altered by IPTW? Please clarify.

o Related to the previous point, I don’t understand the results provided in the figures. It looks like the authors altered the sample size, while this was not stated in the method section? Or am I misunderstanding what the X-axis is the figures? Please clarify

MINOR COMMENTS

- Page 2, line 14: the authors state that conditional analyses target at the individual level. This is only true if the covariates included in the model fully describe individual characteristics. I would rather say that conditional effects apply to specific strata defined by the vector of covariate included the model;

- Figures are too small and really difficult to read

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

Quality of written English: Needs some language corrections before being published

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

Declaration of competing interests:

I declare that I have no competing interests