Random effects modelling vs logistic regression for the inclusion of surgeon covariates in propensity scores for medical device epidemiology: a simulation study

Authors: Mike Du, M Sanni Ali, Victoria Strauss, and Daniel Prieto-Alhambra

Background:
Surgeon covariates like experience or volume can be relevant confounders in medical device epidemiology. This creates challenges in propensity score (PS) estimation due to the multi-level nature of the resulting data, with a variable number of patients operated by a same surgeon.

Objective:
To estimate the performance (bias and accuracy) of random effect modelling (REM) compared to logistic regression (LR) for PS estimation in different surgeon-level effect scenarios.

Methods:
Treatment choice and outcome effect for 10,000 patients were generated using Monte Carlo simulations with multi-level mechanisms (patients nested within surgeons). Five patient confounders, one instrumental variable and one risk factor were generated, all binary. True treatment effect was fixed at Odds ratio (OR) 1.5. A surgeon-level confounder was generated with Pois(2), with an OR = 2 association with treatment, and an OR ranging from 1.5 to 5 for effect on outcome. Two different PS were estimated using: LR including both patient- and surgeon-level confounders; and REM including patient confounders and with surgeon effect as cluster. Inverse probability weighting

(IPW) with non-parametric marginal estimator was used for treatment effect estimation. The resulting treatment effects were compared to the true effect, and % bias and mean square errors (MSE) estimated.

Results:
LR-based PS result in lower bias and MSE than REM-based ones in all simulated scenarios. For example for an OR of 1.5 for surgeon confounding, % bias and MSE was 9.2% and 0.006 for LR compare to 16.4% and 0.009 for REM. Bias and MSE increased in scenarios with stronger surgeon effect on outcome for both LR and REM PS models.

Conclusion:
Using REM for propensity score does not reduce bias compare to plain LR-based propensity scores including surgeon-level confounders. More research is needed on alternative solutions for multi-level PS estimation, and on guidance for their use in medical device epidemiology.