This is a repository copy of *Propensity score stratification and inverse probability weighting compared to surgical randomized clinical trial results in device epidemiology*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/151144/

Version: Accepted Version

TITLE: Propensity score stratification and inverse probability weighting compared to surgical randomised clinical trial results in device epidemiology

V Y Strauss, A Prats-Uribe, K Berencsi, A Carr, A Judge, A Silman, DW Murray, NK Arden, I Petersen, D J Beard, JM Wilkinson, I Douglas, JM Valderas, S Lamb, and D Prieto-Alhambra

Background: Propensity scores are popular in pharmacoepidemiology. However, there is little data on their performance in the study of devices, where confounding is related to more complex mechanisms including not only patient but also surgeon characteristics.

Objectives: to test the performance of propensity score stratification and inverse probability weighting (IPW) to emulate an ongoing surgical RCT comparing partial to total knee replacement, the Total Or Partial Knee Arthroplasty Trial (TOPKAT).

Methods: Patients who underwent either a partial or total knee replacement were identified in the UK National Joint Registry linked to hospital inpatients data (Hospital Episode Statistics England) and patient reported outcome measures (PROMs). TOPKAT eligibility criteria were applied. Logistic regression was used to calculate PS for partial knee replacement using 28 covariates including demographics, preoperative PROMs, comorbidity, and procedures within 3 years before surgery. We compared different methods to the trial results: 1) stratification based on the entire cohort's PS (PSS_{cohort}); 2) stratification based on the PS of patients with partial knee replacement ($PSS_{exposure}$), and 3) IPW. Linear regressions were used to derive average treatment effect (ATE) estimates of difference in postoperative OKS between treatment groups with adjustment for imbalanced covariates (absolute standardized mean differences(ASMD) >0.1). A chi2 test was applied to test for significance differences between TOPKAT findings and each of the proposed methods. Tau2 was used to quantify between study variances.

Results: In total, 355 and 33,982 partial and total knee replacement, respectively, were analysed. $PSS_{exposure}$ resulted in excellent covariate balance (all ASMD<0.1), while many covariates remained imbalanced in IPW and PSS_{cohort}. Postoperative OKS average differences were 1.2 (95%CI: 0.2, 2.1), 1.2(0.1, 2.2) and 1.2 (0.1, 2.2) in favour of partial knee replacement in IPW, PSS_{cohort}, $PSS_{exposure,}$ respectively, compared to 1.8 (0.2, 3.4) in

TOPKAT. All methods obtained Chi2 P>0.05 and small tau2 (<0.0001), suggesting no significant difference in ATE estimates compared to the TOPKAT RCT findings.

Conclusions: PSS$_{exposure}$ obtained better covariate balance than IPW and PSS$_{cohort}$, though all methods resulted in similar treatment effect estimates, all of them comparable to the TOPKAT RCT treatment effect estimate.