AUTHORS: D Prieto-Alhambra, J Weaver, H Morgan-Stewart, A Sena, E Burn, and P Ryan

2,350 chars/max 2,350 (excluding spaces but including Title and Abstract Body)

TITLE: Negative control outcomes to control for unmeasured confounding in device safety research: a clinical example

Background: Negative control outcomes have been proposed to adjust for uncontrolled confounding in drug safety research. There is seldom experience on their use in medical device epidemiology.

Objectives: We aimed to demonstrate the use of negative control outcome calibration after propensity score matching. We tested this in the study of post-operative infections following partial (PKR) compared to total knee replacement (TKR).

Methods:
-Design: Propensity score matched multi-database parallel cohorts.
-Setting and participants: Data was obtained from 4 US claims (IBM MarketScan® Commercial Database (CCAE), IBM MarketScan® Medicare Supplemental Database (MDCR), Optum ® de-identified Clinformatics ® Datamart, Extended - Date of Death (Optum) and PharMetrics) databases. All these were mapped to the OMOP common data model.
Subjects undergoing PKR or TKR in any of these databases from 2005 onwards at age 40 or older were included.
-Exposures: PKR and TKR were compared.
-Outcomes: post-operative infections of the surgical area or the prosthesis within 60 days post-surgery were considered.
-Statistical analysis: Propensity score matching with a 0.02 SD caliper width was used to minimize confounding. Cox regression models were fitted to calculate Hazard Ratio (HR) and 95% Confidence Intervals [95CI] for infection according to device type, using TKR as the reference group. A list of 39 negative control outcomes was pre-specified and results used to calibrate the obtained HRs. All analyses were run separately for each of the databases [Schuemie MJ et al. Stats Med 2014] using a common R package.

Results: Propensity score matching worked to reduce confounding for all observed variables below the pre-specified threshold of a standardized difference of <0.1. In CCAE, 96/7,779 (1.23%) PKR and 845/58,290 (1.45%) TKR subjects had the outcome of interest; compared to 58/4,093 (1.42%) PKR and 613/34,836 (1.76%) TKR in MDCR; 71/5,750 (1.23%) PKR and 795/43,612 (1.82%) TKR in Optum; and 141/12,777 (1.10%) PKR and 1,638/98,516 (1.66%) in PharMetrics. Cox models in the propensity-matched cohorts suggested a consistent reduction in risk of infection with PKR, with HR [95CI] ranging from 0.66 [0.55-0.78] in PharMetrics to 0.85 [0.68-1.05] in CCAE. Calibration after negative

control outcomes attenuated this and resulted in non-significant associations in any of the 4 databases, with HR ranging from 0.73 [0.45-1.24] in PharMetrics to 1.04 [0.77-1.43] in CCAE.

Conclusions: Calibration for negative control outcomes was useful to adjust for residual confounding (for unobserved variables) after propensity score matching.