Validity of conclusions from observational studies depends largely on how studies are designed and analyzed. Investigators must make many decisions, including: patient selection; definitions of baseline, covariates, exposure, follow-up and outcomes; and statistical methods. Extensive testing of analytic tools by the Observational Medical Outcomes Partnership (OMOP) across databases shows that results are highly sensitive to these decisions when screening for drug safety.

Objectives

To explore the impact of design and analysis decisions on covariates balance and risk estimates using data of electronic health databases for active surveillance of marketed drugs.

Methods

Using the Mini-Sentinel protocol for the Active Surveillance of acute myocardial infarction (AMI) in Association with Use of Anti-Diabetic Agents as a template, we applied eligibility, covariate and AMI definitions as in the protocol to the extent possible but comparing metformin (MET) vs second-generation sulfonylureas (SU) in AMI risk. We added the exclusion of first generation SU. We explored various options in the design and analysis.

We defined cohorts of new users of Metformin (M) and second-generation Sulfonylureas (S) between 1/1/2005 and 12/31/2009, baseline covariates and AMI events using 3 definitions of washout and baseline periods (days): (washout 365, baseline 183); (washout 365, baseline 365); and (washout 183, baseline 183).

We assessed covariate balance in quarterly analyses with propensity score (PS) matching and stratification methods applied to cohorts with different washout and baseline periods. The balance metric used for each covariate is the standardized mean difference:

\[
d = \frac{\bar{x}_M - \bar{x}_S}{\sqrt{(s^2_M + s^2_S)/2}}
\]

Risk estimates and confidence intervals were obtained from Cox proportional hazard regression.

Results

- Extending washout from 183 to 365 days decreased new users from 27,420 to 25,675.
- A 365-day baseline identified 78 more patients with possible gestational diabetes and 2% more smokers than a 183-day baseline.
- Prevalence of prior CVD was underestimated by 13% using 183-day baseline than using 365-day baseline, delaying the start of surveillance by 6 months (Figure 1).

- Matching all cumulative patients tended to produce smaller absolute standardized mean differences (Figure 1) in covariates than matching new patients only each quarter, with decreasing variability (Figure 2).

- For analyses with 365 days for both washout and baseline, all hazard ratio (HR) CIs contained 1 (Figure 3).

- With either washout or baseline 183 days, HRs tended to ‘signal’, more so with matching only new patients each quarter.

In PS stratification, PS trimming improved balance whereas stratifying PS further into deciles did not. HRs were similar to those from PS matching.

Conclusions

- Longer durations of washout and baseline may enhance the classification of new users and better capture comorbidities but at reduced sample size.
- Matching all cumulative patients offered better balance than matching only new patients each quarter.
- PS matching achieved better balance than PS stratification but with only half the sample size.

References