Calibrating The Strength Of Association Of Drug-outcome Pairs Using The Empirical Null Derived From Known Non-associations And Known Positive Associations Identified In The Observational Medical Outcomes Partnership (OMOP)

Xiaochun Li, PhD[1, 2], Susan Ofner, MS[1], Changyu Shen PhD[1, 2], Joyce Zhan MS[2], Zhiwen Liu PhD[4], Marc Rosenman MD[1, 3], Nancy Santanello MD, MS[4]

[1]Indiana University School of Medicine, [2]Indiana University Richard M. Fairbanks School of Public Health;[3]Regenstrief Institute, [4]Epidemiology, Merck Sharp & Dohme

INDIANA UNIVERSITY SCHOOL OF MEDICINE

Regenstrief Medical Informatics
The Source for Medical Informatics

INTRODUCTION

Electronic medical databases can be used to investigate drug-outcome associations (drug safety). Often relative risks (RR) are calculated using appropriate methods and adjusting for a list of putative confounders. Under the null hypothesis of no association, we expect an RR to be close to 1 and quantify the strength of association by gauging how far the RR estimate deviates from 1. However, EMR is created for routine care and administration, not for research; it may have hidden biases from failure or incomplete capture of exposure, covariates and outcome. This innate bias may lead to RR's null value not centered at 1. If not taken into account, risk assessment will be either missed or exaggerated (both concerns public health). By using negative control exposure-outcome pairs, Schuemie et al (Statistics In Medicine, 2013) quantify the systematic error that traditional significance testing fails to take into account and produce calibrated p-values, allowing for more accurate interpretation of the results of observational studies. As a fingerprinting exercise, we used the same set of OMOP controls to gain a better understanding on the direction and the magnitude of bias in a pharmaco-epidemiology study in the Regenstrief database (RI CDM).

DATA SOURCE

The Indiana Network for Patient Care (INPC) is a heath information exchange-based clinical repository containing medical records on over 11 million patients throughout the state of Indiana. The Regenstrief Common Data Model (CDM) is a derivation of the INPC containing coded medication, diagnosis, and observation data on 2.2 million patients between 2004 and 2009. These data were architected specifically for research on adverse drug reactions through collaboration with the Observational Medical Outcomes Partnership (OMOP).

METHODS

We built our exploration on the OMOP set of 102 control drugs for acute myocardial infarction (AMI): 36 drugs with positive associations and 66 drugs with no associations with AMI, established through literature review and expert panel assessment . Using the Regenstrief CDM, for each drug we identified patients with 1) an event for drug exposure, and 2) an outcome of AMI (OMOP definition 3). We implemented the self-controlled case series (SCCS) algorithm with age adjustment using 5-year intervals to estimate the incidence density ratio (IDR) of AMI for exposure versus non-exposure to each of the control drugs.

Each drug exposure was analyzed using 12 combinations of analytic options of 6 "at-risk" intervals by 2 event types (first or recurrent) in the SCCS analysis. Surveillance window (sw) defines the period of time that a patient is inferred to be 'at-risk'of AMI based on drug prescription, dispensing, or administration. The 6 options are -30, 0, 30, 60, 90 and 120, corresponding to 6 risk intervals, from each exposure start date to 30 days post the start of exposure or to 0, 30, 60, 90 or 120 days post the end of exposure respectively. For summary and plotting, we used only those controls (drugs) that have at least 10 unique patients. Positive controls were used to construct ROC curves.

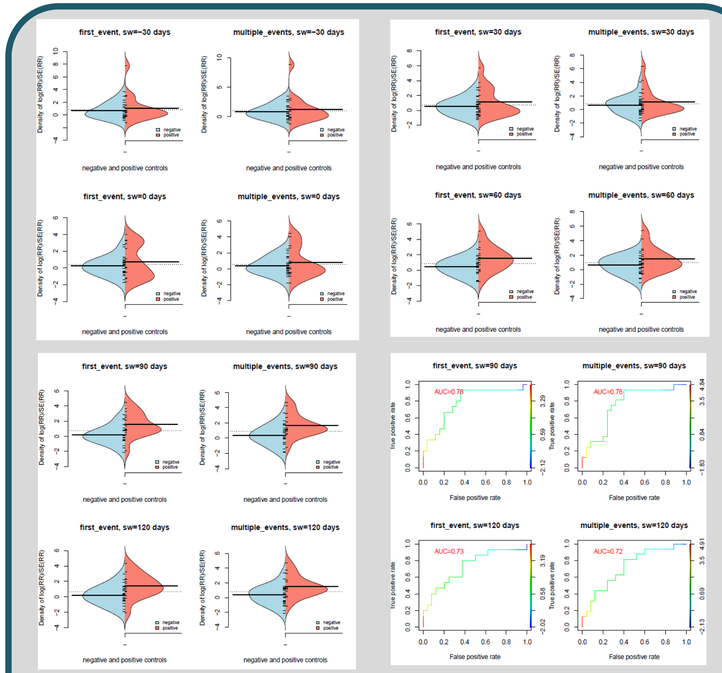
RESULTS

Table 1

Surveillance Window (days)	Event Type	N_{-}	N^{-}_{+}
-30	First	26	15
	Multiple	26	15
0	First	25	15
	Multiple	24	14
30	First	25	15
	Multiple	25	17
60	First	24	15
	Multiple	24	16
90	First	25	15
	Multiple	25	16
120	First	24	15
	Multiple	25	16

Table 1shows the numbers of positive and negative controls for the twelve combinations of analytic options in SCCS.

Due to either insufficient sample sizes or the failure in algorithm convergence, we have results on fewer drugs than drugs in the OMOP control set.



The first 3 plots are distributions for +/- controls. AUCs for surveillance windows (sw) -30, 0,30 are 49%-55%, and 71% (incident) and 64% (recurrent) for sw 60, maximizing at sw=90 days. We analyzed with sw=120 post hoc to test if AUC appeared to be a function of sw.

CONCLUSION

1. The at-risk window of from the index exposure date to 90 days post the end date of exposure gives the best separation of the negatives and the positives.
2. Using the first AMI in the analysis discriminates negative and positive controls better than using all recurrent AMIs.
3. Although the AUC tends to increase as exposure eras are extended, it plateaus at 90 days. The AUC decreases if exposure eras are further extended by 120 days.
4. Using the 90 days extension to exposure eras and the first AMI event gives the best result with an AUC 78%.