Nina Joyce
Brown University School of Public Health
Department of Epidemiology
121 South Main St.
Providence, RI 02903
Nina_joyce@brown.edu
(339) 221-0866

Comparison of a propensity score matched retrospective cohort to a nested case-control study design for post-market surveillance of rare drug exposures in pediatric populations

Proposal application to the
Innovation in Medical Evidence Development and Surveillance (IMEDS)

Nina Joyce
February 5, 2015

Research Objectives and Aims:

As advances in computing power and longitudinal data methods in observational settings have advanced, so too has the use of large administrative health claims databases (AHCDs), such as the Marketscan data commercial claims data, in public health research. Their size makes AHCDs useful for the study of rare exposures and outcomes, and their method of data collection, passive rather than active, avoids the recall bias often associated with retrospective interview studies.[1] These characteristics are especially advantageous for post-market surveillance in pediatric populations, where randomized controlled trials are less frequent.

The most common method of analysis in these databases is the retrospective cohort study, in which a patient's exposure to a drug is identified and the patient is then followed over time until they have the outcome or are censored.[2,3] In this scenario, Cox proportional-hazard models are often used to compare the hazards of the exposed population to the un-exposed while controlling for a variety of covariates measured from the database. The benefit of the retrospective cohort design is that it takes advantage of information from all patients in the sample. Recently, however, the nested case-control (NCC) design has gained popularity as a means of analyzing these data due to its computational efficiency and the relative ease of including time-varying covariates by sampling controls as a function of their available person-time.[2,3] Historically, case-control designs have been limited by the difficulty in finding an unbiased sampling frame for the controls.[4] However, the NCC design samples from the same cohort as the retrospective cohort design, and the size of AHCDs, sometimes in the tens of millions, means that finding multiple controls is not often a problem.

The objective of this study is to compare the use of the retrospective cohort to the NCC for the study of side effects associated with a rare drug exposure in a pediatric population. To illustrate this comparison we will use the example of the association between statin use and diabetes in children ages 8 to 20. Several clinical trials and observational studies have found a causal effect of statin use on the development of incident diabetes in adults.[5,6] However the relationship has never been studied in a pediatric population where the cost-benefit ratio may not be as favorable. To achieve this objective we propose the following aims:

1. Estimate the risk of incident diabetes associated with statin use in a population of children ages 8 to 20 using a propensity score matched retrospective cohort study design.
2. Estimate the risk of incident diabetes associated with statin use in a population of children ages 8 to 20 using a matched nested case-control study design.
3. Compare estimates from the cohort study in aim 1 to those from the NCC in aim 2.

Scope/Proposed Approach

Aim 1: Estimate the risk of incident diabetes associated with statin use in a population of children ages 8 to 20 using a propensity score matched retrospective matched cohort study design.

We propose to use a propensity score matched cohort to estimate the effect of statin use on development of incident diabetes using a Cox proportional hazard model. Incident diabetes will be defined as the first of two outpatient claims in a 12 month period or a single inpatient claim for diabetes (ICD9 code 252.x). Statin use will be identified from prescription pharmacy claims and separate models will be specified for new users (no recorded statin dispensings in the past 6 months) and prevalent users. Year specific propensity scores will be calculated using logistic regression and will include all known confounders or their proxy measures available in the data. We will use a 5-to-1 greedy matching algorithm with replacement to select up to 10 untreated matches within a 0.01 caliper of the treated

patient's propensity score for each year (see Seeger 2007) for a more detailed description of the year-specific propensity score). [7] As a sensitivity analysis we will construct a separate propensity score on the subsample of patients with available lab values in the data to test the robustness of our propensity score to the inclusion of lab values, which are not as always available in AHCD's.

One of the primary concerns in observational studies of drug safety is the potential for confounding by indication (COI), in which characteristics associated with a physician's decision to prescribe a drug or a patient's decision to use a drug is related to the outcome of interest. The propensity score, a measure of an individual's probability ("propensity") of being exposed to a drug, is one proposed method for dealing with COI. Matching on the propensity score results in a cohort of patients that is balanced with respect to the covariates included in the propensity score. Further, if all confounders are included, it provides a direct estimate of the average treatment affect among the treated.[3,8]

The ability of the propensity score to control for COI, however, depends on the availability of data on confounders of the treatment-outcome relationship. AHCDs contain a wealth of clinical and demographic information, making them useful for propensity score analyses in this respect. Yet, data in AHCDs are collected for billing rather than research purposes, and as such, patterns in the data often reflect incentives related to reimbursement rather than treatment or disease, which can limit their ability to accurately measure confounders. Several studies, however, have shown that inclusion of proxies for confounders in the propensity score model can work to control for confounding when actual measures are not present.[9,10] Thus, while AHCDs may not contain information on all confounders, it may be possible to adjust for confounders by including proxy measures that are measured in AHCDs.

Aim 2: Estimate the risk of incident diabetes associated with statin use in a population of children ages 8 to 20 using a matched Nested case-control study design

To estimate the effect of statin use on development of incident diabetes in children ages 8 to 20, we will match each new case of diabetes to a maximum of 10 controls sampled from the population at risk at the time of diagnosis. Incident diabetes and statin use will be defined according to the same criteria described in the first aim, and separate models will also be specified for new and prevalent statin users. In addition to matching on time, we will match on demographics, clinical comorbidities and measures of health care utilization. We will use conditional logistic regression to calculate effect estimates, regressing the binary measure of statin use (yes/no), as well as covariates not used to match cases to controls, on the binary outcome of diabetes (yes/no).

Case control studies have historically been limited by the need to identify an unbiased sampling frame from which to select controls.[4] The NCC, however, is not subject to this limitation, as the sampling frame is the same cohort which gave rise to the cases. In AHCDs this cohort can be defined as all patients for whom data is available, or can reflect an event internal to the data, such as all patients with a specific diagnosis, or external to the data, such as all patients in the database on a specific date. The NCC takes advantage of the AHCD data structure by sampling a control, or multiple controls, from the available risk set whenever a case occurs. In this way, the NCC accounts for censoring, similar to the cox-model, by sampling controls proportional to their available person-time. Thus, patients with less person-time have a lower probability of being sampled as a control while patients with more person time have a greater probability.[11]

Aim 3: Compare estimates from the cohort study in aim 1 to those from the NCC in aim 2

In simulation studies the retrospective cohort has been shown to result in less bias and greater precision than the NCC.[3] Estimates from a NCC that are similar to those from the retrospective cohort study would suggest that the same study could be done more efficiently using a NCC.

Impact: One of the strengths of the NCC, relative to the cox model, is the computational ease with which it handles time-varying covariates. Because controls are sampled at the same time as the case, the covariates for both the case and control only need to be evaluated at a single point in time. In contrast, the cox proportional-hazards model requires the investigator to structure the data to account for each change in a time-varying covariate, which can be a monumental task when including the multitude of time-varying covariates available in AHCDs.[12] Results suggesting that the NCC can be used in place of the retrospective cohort study design would allow investigators to more efficiently identify safety risks associated with rare pharmaceutical drug use. In the case of pediatric populations where exposures and outcomes are rare, the use of the NCC would be particularly beneficial.

Experience: Nina Joyce will be the only person requiring access to the data. Her dissertation committee will review output that she generates from her analyses, but will not require individual access to the data.

Timeline: We request access to the data for 18 months. We anticipate completing the analysis within 12 months and then accessing the data for manuscript revision in the subsequent six months. Thus we request start and end dates of March 1st, 2015 and September 30th, 2016 respectively.

References:

1. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *Journal of clinical epidemiology.* Apr 2005;58(4):323-337.
2. Essebag V, Platt RW, Abrahamowicz M, Pilote L. Comparison of nested case-control and survival analysis methodologies for analysis of time-dependent exposure. *BMC medical research methodology.* Jan 25 2005;5(1):5.
3. Austin PC, Anderson GM, Cigsar C, Gruneir A. Comparing the cohort design and the nested case-control design in the presence of both time-invariant and time-dependent treatment and competing risks: bias and precision. *Pharmacoepidemiol Drug Saf.* Jul 2012;21(7):714-724.
4. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. I. Principles. *American journal of epidemiology.* May 1 1992;135(9):1019-1028.
5. Culver AL, Ockene IS, Balasubramanian R, et al. Statin use and risk of diabetes mellitus in postmenopausal women in the Women's Health Initiative. *Archives of internal medicine.* Jan 23 2012;172(2):144-152.
6. Sattar N, Preiss D, Murray HM, et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. *Lancet.* Feb 27 2010;375(9716):735-742.
7. Seeger JD, Kurth T, Walker AM. Use of propensity score technique to account for exposure-related covariates: an example and lesson. *Medical care.* Oct 2007;45(10 Supl 2):S143-148.
8. Austin PC. The performance of different propensity-score methods for estimating differences in proportions (risk differences or absolute risk reductions) in observational studies. *Statistics in medicine.* Sep 10 2010;29(20):2137-2148.
9. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology.* Jul 2009;20(4):512-522.
10. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. *American journal of epidemiology.* Jun 15 2006;163(12):1149-1156.
11. Vandenbroucke JP, Pearce N. Case-control studies: basic concepts. *Int J Epidemiol.* Oct 2012;41(5):1480-1489.
12. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. *Annual review of public health.* 1999;20:145-157.