Editorial

Using the Market Scan Database to Conduct Retrospective Cohort Studies on Dermatologic Treatments and Conditions: Benefits and Pitfalls

Huang KE and Davis SA*

Department of Dermatology, Wake Forest School of Medicine, USA

Claims databases serve as invaluable tools in dermatologic research. However, their worth can be jeopardized by poor study designs. Here, we describe one claims database as an example of claims database structure, detail benefits of using a claim database for epidemiologic studies, and present methods that can be used to address some major biases that may arise when analyzing such data.

An example: The MarketScan® research database (Truven Health Analytics, Ann Arbor, Mich.) longitudinally collects information from medical encounters made by a sample of patients in the United States [1]. Data are available for over 170 million patients who have been sampled since 1995. Patients are divided by type of insurance carried (employer purchased, Medicaid/ Other Public, Medicare, individual, uninsured). Types of data available, which can be used for research, include patient demographics, occurrence of inpatient and outpatient visits, diagnoses rendered and procedures performed at the visits, dispensing of medication, length of enrollment, lab test results, and payment records for healthcare services or medication.

Where is the value?

Claims databases offer access to large samples of patients, which can facilitate investigations of rare diseases or events. In many cases, clinical trials of the same magnitude and detail would be financially and logistically infeasible as well as sometimes unethical. As all medical encounters and lab results are included for each patient, a wide array of diseases and covariates can be studied. Finally, the design of the database enables long-term surveillance of medication use, which can be beneficial for pharmacoepidemiologic studies.

What can go wrong? How to minimize the problems?

While there are clear benefits for epidemiologic studies in dermatology, problems can arise if the inherent flaws of claims databases are ignored.

Misclassification of variables: Misclassification is a mislabeling of a patient's status. For example, in a claims database study of acne, eligible patients may be identified with a diagnosis of acne. Any patients that do have acne but never received a diagnosis from their physician will be misclassified as not having acne. Alternatively, some patients may be misdiagnosed as having acne and would be misclassified as having acne.

There can be several different results from misclassification. If misclassification of a dichotomous exposure is not related to the outcome being researched, then the estimate of effect will be attenuated toward the null. However, if there is a correlation between the outcome value and the rate of misclassification of an exposure, or if there are more than two levels of an exposure (e.g. no disease, mild disease, and severe disease) then there can be biases [2,3]. To measure or adjust for error, a validation study, which consists of a sample of the cohort, can be constructed to estimate the rates of misclassification (or measurement error for continuous variables), [3] which can then be used in misclassification models. However, a validation study is not always feasible. To minimize false positives, cases can be restricted to patients who received multiple diagnoses of a disease at different visits or who were diagnosed and filled a disease-related medication [4].

Missing data: When different patients do not receive the same laboratory tests or types of medical visits, some patients will have missing data for covariates of interest. One method to handle this is a complete-case analysis, in which only patients with no missing data are included in the analysis. If many subjects are missing data, then the final analysis cohort may be too small [3]. Additionally, if the data are missing not due to chance, then biases can be introduced. For example, if ordering LDL levels of cholesterol was related to the obesity of the patient and obesity was an outcome of interest, then the analysis would likely be biased if subjects with missing LDL data were excluded.

An alternate method, which could be presented as a secondary analysis, is imputation of the missing variables. In this case, values for the variables with missing data are estimated based on other recorded variables. For these methods, the large assumption is that the missingness of the data can be explained by data that are recorded [3,5]. Using the above example, if the LDL cholesterol data were missing contingent only on obesity status, then an imputation model may be valid. If values were

Davis et al. (2013)
Email: scdavis@wakehealth.edu

◉SciMedCentral

missing for other reasons that are not included (e.g. smoking status, patient's willingness to have blood drawn), then the imputation models will not be accurate. As both of these methods can introduce further bias, thoughtful deliberation should be taken when deciding how to handle the missing data.

Confounding by indication: For drug comparator studies, individuals using one treatment commonly are compared to individuals who did not receive the treatment. Confounding can arise when dissimilar subjects are compared. For example, a patient with psoriasis who is prescribed a mid-potency topical corticosteroid will likely not have the same disease severity as a psoriasis patient using a biologic. If disease severity is positively related to the outcome of interest (e.g. cardiovascular disease) then the estimate of effect will be biased upwards. Stratification of the cohorts into groups with similar disease severity can help minimize this concern [6]. Other methods include using a matching design or restricting the cohort to similar populations [3]. As these methods unlikely will completely correct for such confounding, researchers should also acknowledge this limitation of their observational research [7].

Unmeasured confounders: There may always be some level of unmeasured confounding present in a study. To assess the magnitude of its effect, a sensitivity analysis can be performed [8]. Results from this analysis can identify how robust the study's findings are in light of potentially unidentified confounders. An external validation study or a supplementary data collection, when feasible, can also be used. From these additional data sources can be used to adjust the estimates of effect [3,9]. Sometimes case crossover studies, which use cases acting as their own controls, can be ideal for controlling time non-varying unmeasured confounders [10,11].

Despite these and other potential biases, the MarketScan and other claims databases are some of the best tools available to answer dermatologic research questions when clinical trials are not feasible.

CONFLICT OF INTEREST

REFERENCES

1. Hansen L, Chang S. White Paper - Health Research Data for the Real World: The MarketScan Databases.Truven Health Analytics. 2012.

2. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value?. Am J Epidemiol. 1990; 132: 746-8.

3. Rothman K, Greenland S, Lash T. Modern Epidemiology, 3rd ed. Philadelphia: Lippincott Williams & Wilkins. 2008.

4. Icen M, Crowson CS, McEvoy MT, Gabriel SE, Maradit Kremers H. Potential misclassification of patients with psoriasis in electronic databases. J Am Acad Dermatol. 2008; 59: 981-5.

5. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009; 338: 2393.

6. Walker AM. Confounding by indication. Epidemiology. 1996; 7: 335-336.

7. Shapiro S. Confounding by indication?. Epidemiology. 1997; 8: 110-111.

8. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. Pharmacoepidemiol Drug Saf. 2006; 15: 291-303.

9. Stürmer T, Glynn RJ, Rothman KJ, Avorn J, Schneeweiss S. Adjustments for unmeasured confounders in pharmacoepidemiologic database studies using external information. Med Care. 2007; 45: 158-65.

10. Maclure M, Fireman B, Nelson JC, Hua W, Shoaibi A, et al. When should case-only designs be used for safety monitoring of medical products?. Pharmacoepidemiol Drug Saf. 2012; 1: 50-61.

11. Maclure M. 'Why me?' versus 'why now?'--differences between operational hypotheses in case-control versus case-crossover studies. Pharmacoepidemiol Drug Saf. 2007; 16: 850-3.