Received: 7 January 2018 Revised: 5 October 2018 Accepted: 11 October 2018
DOI: 10.1002/pds.4694

Pharmacoepidemiology and Drug Safety's special issue on validation studies

Danielle S. Chun | **Jennifer L. Lund** | **Til Stürmer**

Epidemiology, University of North Carolina at Chapel Hill Gillings School of Global Public Health, Chapel Hill, North Carolina, USA

Correspondence
D. S. Chun, Epidemiology, University of North Carolina at Chapel Hill Gillings School of Global Public Health, Chapel Hill, NC, USA.
Email: danielle.chun@unc.edu

1 | INTRODUCTION

Administrative claims and other routinely collected data provide the foundation for many drug utilization, safety, and effectiveness studies. These databases provide a rich source of timely health care information on large, well-defined populations. Yet information contained in these databases is generally captured using standardized systems, summarizing complex medical histories, clinical diagnoses, and services and therapies provided to patients. Thus, carefully designed validation studies that evaluate the accuracy of coded algorithms to identify health-related exposures, outcomes, and covariates against a reference standard are an essential component for demonstrating the validity of their use for research purposes.

Misclassification can occur when coded algorithms inaccurately reflect the true value of a given exposure, outcome, or covariate over a specified time period.[1] Drug exposures obtained through prescription fills and other procedure codes are generally well captured in administrative data and are often viewed as the reference standard when compared with self-reported information or records, physician-ordered prescriptions, or surveillance data.[1] In addition, the degree of bias because of misclassification of a confounder is typically bounded by its confounding effect and is often of secondary concern.[2]

Because many health care databases lack detailed clinical and pathologic information, misclassification of outcome or disease incidence remains a major challenge in pharmacoepidemiologic research. For instance, prevalent disease at the outset of a study may be misclassified as absent, and thus, the first coded observation during follow-up could be misclassified as incident disease. True incident disease can be missed due to low algorithm sensitivity (ie, high false-negative rate), resulting in artificially low estimates of disease incidence. True absence of incident disease may be miscoded as incident disease (ie, high false-positive rate) when codes are used to "rule out" suspected disease, leading to overestimates of disease incidence.[1] In studies evaluating the effects of drug exposures on intended and unintended events, outcome misclassification can lead to biased treatment effect estimates[1,3] The direction of this bias depends in part on whether the outcome misclassification is differential with respect to the drug exposure.[3] In the setting of nondifferential outcome misclassification with perfect specificity, relative risk estimates remain unbiased, while absolute risks estimates do not.[1,3] When the extent of outcome misclassification depends upon on exposure status, treatment effect estimates can be biased either towards or away from the null.[3]

Pharmacoepidemiologists often develop coded algorithms by combining diagnosis codes (and their position on the claim) on unique service dates, prescription fills, claims from an inpatient setting, and/or specific procedures to improve the specificity of the outcome definition. To ensure study validity, algorithms should be validated against a reference standard and report measures of validity to assess their accuracy. Researchers should first consider the strengths and limitations of available reference standards, as data accuracy may depend on how information is obtained (eg, via clinical assessment or linkage with a cancer registry).[4] Moreover, researchers need to question the transportability of a given algorithm to external settings as each algorithm is developed for a specific study population, database, time period, and research question. Thus, validation studies should report the characteristics of the data source (including calendar years), study population (eg, older patients with diabetes), and measures of validity used to quantify the extent of misclassification. Comprehensive reporting of measures of validity (ie, specificity, sensitivity, positive predictive value (PPV), and negative predictive value (NPV)) for coded algorithms is especially useful for researchers applying these algorithms in external settings who often must prioritize certain measures of validity depending on their study goals.[5] While evaluating the disease status of algorithm-defined negative cases (ie, false negatives and true negatives) can be challenging, this information is required for reporting algorithm sensitivity and specificity. This information is also useful in quantifying the impact of misclassification on study results through quantitative bias analysis.[6]

There is a growing need to develop and validate new coded algorithms and to investigate their transportability to various populations, health care settings, coding systems, and calendar time periods.

Algorithms may become outdated due to temporal changes in coding systems within health care databases. A recent study compared the validity of coding for several clinical conditions in administrative hospital discharge data in Canada during the transition from International Classification of Diseases (ICD), Ninth Revision, Canada (ICD-9-CA) coding to ICD-10-CA coding, which occurred from 2001 to 2004.[7] Overall, findings showed high validity, yet for some conditions, including myocardial infarction, hypertension, and depression, the sensitivity in ICD-10-CA was significantly lower than that in ICD-9-CA. Thus, updated validation studies can provide data needed to ensure adequate capture of health outcomes under new disease classification systems. This is particularly relevant in the United States as many of the claims and electronic health record data used for research now include codes based on both ICD-9 and ICD-10.

Algorithms may also require reevaluation because of changes in how health care is delivered and reimbursed. In 2004, a validation study by Kiyota et al assessed the PPV of acute myocardial infarction (AMI) discharge diagnoses codes.[8] The study population included Medicare beneficiaries in Pennsylvania enrolled in that state's Pharmaceutical Assistance Contract for the Elderly (PACE), a prescription drug program for low-income elderly individuals, in 1999, 2000, or both years. A claims-based definition was developed using ICD-9 diagnosis codes, diagnosis-related groups, and included a hospitalization episode lasting at least 3 days.[8] The reported PPV for this algorithm was 95.4% (95% CI, 94.3%-96.4%). However, the duration of hospitalization for AMI has decreased considerably over time.[9] To address these changes, Brouwer et al. reassessed and validated claims-based algorithms for AMI in a clinical cohort of HIV patients with Medicaid claims data between 2002 and 2008. The authors reported a lower algorithm PPV with a length of stay lasting 3 or more days, 47.8% (95% CI: 26.8%-69.4%), compared with any length of stay, 62.5% (95% CI: 35.4% – 84.8%).[10] Despite the fact that differences in PPV might be partially explained by underlying differences in study populations, periodic reevaluation of outcome algorithms to investigate the impact of changes to health care delivery is warranted.

Algorithms should also be routinely reassessed in different patient populations, data sources, and time periods. Populations may have different baseline risks for health outcomes and health care data may be collected differently across databases, leading to issues with algorithm transportability. In 2009, another landmark validation study by Setoguchi et al evaluated the accuracy of claims-based definitions to identify incident cancers.[11] The study population included Medicare beneficiaries in Pennsylvania enrolled in PACE between 1997 and 2000. The algorithms used claims data from 6 months before the start of the study period to remove prevalent cancer cases from the eligibility for incident cancer identification. Across cancer types, these definitions had very high specificity (≥98%). A later validation study by Bronson et al updated the algorithm by Setoguchi et al. using Medicare claims data linked to the Nurses' Health Study (NHS).[12] The cohort included female registered nurses, age 30 to 55 years, living in 11 US states and used a 2-year study period. Bronson et al reported similarly high algorithm specificity (≥98%) for all cancers, yet found differences in algorithm sensitivity for incident breast, colorectal, and lung cancer between the two studies. This may be partially explained by the fact that NHS participants were female health professionals, younger, mostly white, and not a very low income population.

The increasing ability to link data sources provides researchers with opportunities to validate claims-based algorithms using internal reference standards that include key clinical information often unavailable to claims data. A salient example is the Surveillance, Epidemiology, and End Results (SEER)-Medicare data, a linkage between cancer registry and Medicare claims data, which have been used to validate algorithms for breast cancer incidence.[13] Recently, data was linked across Kaiser Permanente Northwest and Northern California electronic health records (EHRs), claims data for members receiving services outside the KP systems, and death data files to assess the validity of ICD codes for identifying opioid-related overdoses.[14] Future directions for validation studies may consider conducting bias analyses and incorporating innovative methods to correct for misclassification and improve the accuracy of case ascertainment such as machine learning approaches and natural language processing. In addition, regression models can be used to estimate the probability of being a confirmed case instead of simply classifying a subject as a case or noncase.[4] Split sample approaches can also be used to reduce the risk of overfitting algorithms to a given study data set.

As the availability of linked health care data for research increases, well-defined algorithms and validation studies for defining specific health events are needed to quantify and mitigate misclassification bias. Further, validation studies must be continuously updated over time and in different settings to ensure continued validity and transportability. In this special issue, several studies address critical issues regarding the validity of outcome algorithms. For example, Ferguson et al assess the accuracy of osteoarthritis identified using Read codes compared with alternative reference standards based on radiologic and clinical criteria using the UK Clinical Practice Research Datalink. Crothers et al highlight issues of transportability to specific populations by comparing the accuracy of ICD-9 diagnosis codes and measures of spirometry to define COPD patients living with and without HIV. Finally, Koram et al highlight the need for tailored algorithms in a comprehensive review of validation studies in the Asia-Pacific Region. This special issue also features innovative approaches to developing coded algorithms and misclassification correction. Beachler et al employ both logistic and LASSO regression modeling to identify cases of early and advanced stage ER+/HER2- breast cancer and Gravel et al develop an approach utilizing conditional validation sampling to correct for potential misclassification of binary health outcomes.

We hope that this issue of *Pharmacoepidemiology and Drug Safety* (PDS) helps to elevate the importance of validation studies for the field of pharmacoepidemiology. We encourage all pharmacoepidemiology researchers to conduct validation studies and submit them for publication in PDS.

ORCID

Danielle S. Chun http://orcid.org/0000-0001-5913-9195
Jennifer L. Lund http://orcid.org/0000-0002-1108-0689
Til Stürmer http://orcid.org/0000-0002-9204-7177

REFERENCES

1. Funk MJ, Landi SN. Misclassification in administrative claims data: quantifying the impact on treatment effect estimates. *Curr Epidemiol Rep*. 2014;1(4):175-185.
2. Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol*. 1980 Oct;112(4):564-569.
3. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
4. Lanes S, Brown JS, Haynes K, Pollack MF, Walker AM. Identifying health outcomes in healthcare databases. *Pharmacoepidemiol Drug Saf*. 2015 Oct;24(10):1009-1016.
5. Chubak J, Pocobelli G, Weiss NS. Tradeoffs between accuracy measures for electronic health care data algorithms. *J Clin Epidemiol*. 2012 Mar;65(3):343-349.
6. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol*. 2014 Dec;43(6):1969-1985.
7. Quan H, Li B, Saunders LD, et al. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res*. 2008 Aug;43(4):1424-1441.
8. Kiyota Y, Schneeweiss S, Glynn RJ, Cannuscio CC, Avorn J, Solomon DH. Accuracy of Medicare claims-based diagnosis of acute myocardial infarction: estimating positive predictive value on the basis of review of hospital records. *Am Heart J*. 2004 Jul;148(1):99-104.
9. Saczynski JS, Lessard D, Spencer FA, et al. Declining length of stay for patients hospitalized with AMI: impact on mortality and readmissions. *Am J Med*. 2010 Nov;123(11):1007-1015.
10. Brouwer ES, Napravnik S, Eron JJ Jr, et al. Validation of Medicaid Claims-based Diagnosis of Myocardial Infarction Using an HIV Clinical Cohort. *Med Care*. 2015 Jun;53(6):e41-e48.
11. Setoguchi S, Solomon DH, Glynn RJ, Cook EF, Levin R, Schneeweiss S. Agreement of diagnosis and its date for hematologic malignancies and solid tumors between medicare claims and cancer registry data. *Cancer Causes Control*. 2007 Jun;18(5):561-569.
12. Bronson MR, Kapadia NS, Austin AM, et al. Leveraging linkage of cohort studies with administrative claims data to identify individuals with Cancer. *Med Care*. 2018 Jan;12.
13. Gold HT, Do HT. Evaluation of three algorithms to identify incident breast cancer in Medicare claims data. *Health Serv Res*. 2007;42(5):2056-2069.
14. Green CA, Perrin NA, Janoff SL, Campbell CI, Chilcoat HD, Coplan PM. Assessing the accuracy of opioid overdose and poisoning codes in diagnostic information from electronic health records, claims data, and death records. *Pharmacoepidemiol Drug Saf*. 2017 May;26(5):509-517.
15. Ferguson R, Prieto Alhambra D, Walker C, et al. Validation of hip osteoarthritis diagnosis recording in the UK clinical practice research datalink. *Pharmacoepidemiol Drug Saf*. 2018.
16. Crothers K, Rodriguez C, Nance R, et al. Accuracy of electronic health record data for the diagnosis of chronic obstructive pulmonary disease in persons living with HIV and uninfected persons. *Pharmacoepidemiol Drug Saf*. 2018.
17. Koram N, Delgado M, Stark J, Setoguchi S, de Luise C. Validation studies of claims data in the Asia-Pacific region: a comprehensive review. *Pharmacoepidemiol Drug Saf*. 2018.
18. Beachler D, de Luise C, Yin R, Gangemi K, Cochetti P, Lanes S. Identifying cases of early and advanced stage ER+/HER2- breast cancer in electronic claims data using predictive modelling. *Pharmacoepidemiol Drug Saf*. 2018.
19. Gravel C, Farrell P, Krewski D. Conditional validation sampling for consistent risk estimation with binary outcome data subject to misclassification. *Pharmacoepidemiol Drug Saf*. 2018.

How to cite this article: Chun DS, Lund JL, Stürmer T. *Pharmacoepidemiology and Drug Safety*'s special issue on validation studies. *Pharmacoepidemiol Drug Saf*. 2019;28:123–125. https://doi.org/10.1002/pds.4694