Use of Health Care Claims Data to Study Patients with Ophthalmologic Conditions

Joshua D. Stein, MD, MS,[1] *Flora Lum, MD,*[2] *Paul P. Lee, MD, JD,*[1] *William L. Rich III, MD,*[3]
Anne L. Coleman, MD, PhD[2,4]

Objective: To describe what information is or is not included in health care claims data, provide an overview of the main advantages and limitations of performing analyses using health care claims data, and offer general guidance on how to report and interpret findings of ophthalmology-related claims data analyses.

Design: Systematic review.

Participants: Not applicable.

Methods: A literature review and synthesis of methods for claims-based data analyses.

Main Outcome Measures: Not applicable.

Results: Some advantages of using claims data for analyses include large, diverse sample sizes, longitudinal follow-up, lack of selection bias, and potential for complex, multivariable modeling. The disadvantages include (a) the inherent limitations of claims data, such as incomplete, inaccurate, or missing data, or the lack of specific billing codes for some conditions; and (b) the inability, in some circumstances, to adequately evaluate the appropriateness of care. In general, reports of claims data analyses should include clear descriptions of the following methodological elements: the data source, the inclusion and exclusion criteria, the specific billing codes used, and the potential confounding factors incorporated in the multivariable models.

Conclusions: The use of claims data for research is expected to increase with the enhanced availability of data from Medicare and other sources. The use of claims data to evaluate resource use and efficiency and to determine the basis for supplementary payment methods for physicians is anticipated. Thus, it will be increasingly important for eye care providers to use accurate and descriptive codes for billing. Adherence to general guidance on the reporting of claims data analyses, as outlined in this article, is important to enhance the credibility and applicability of findings. Guidance on optimal ways to conduct and report ophthalmology-related investigations using claims data will likely continue to evolve as health services researchers refine the metrics to analyze large administrative data sets. *Ophthalmology 2014;121:1134-1141 © 2014 by the American Academy of Ophthalmology.*

Many studies are using claims data and other large administrative health databases to assess ophthalmology-related research questions. Findings from such studies are considered by health policymakers, third-party payers, and other decision makers as they grapple with timely challenges, such as allocating limited resources and finding ways to improve patient outcomes and patient care. Clinicians also are using data to assess their practices and performance compared with their peers and to understand the risk factors and outcomes of their patients. Other uses include estimating adherence to therapy and evaluating utilization.

Interpreting and understanding the generalizability of the findings of these types of studies can be difficult because the methodological rigor of the analyses and the extent of reporting can vary considerably. We provide an overview of the types of information contained in claims data and describe some of the advantages and limitations of using claims data for research purposes.[1] We provide a suggested checklist for authors to use in reporting analyses involving claims or other administrative health data (Table 1), as we aim to improve the overall quality and usefulness of such reports.

Information Contained in Claims Data

Numerous health care claims databases have been used by researchers. According to the International Society of Pharmacoeconomics and Outcomes Research, 382 such databases currently exist, including 153 in the United States alone (http://www.ispor.org/digestofintdb/countrylist.aspx; accessed August 29, 2013). These databases vary from large datasets capturing health care services provided to a nationwide sample of enrollees in Medicare, Medicaid, or managed-care networks, such as the Healthcare Cost and Utilization Project, to those that capture health care services in a specific community, such as Kaiser Permanente Northwest, or focus on patients with certain specific diseases, such as the Framingham Heart Study database. In some countries with a national health system, all health care encounters can be captured systematically and stored in a single repository. However, in the United States, patients can move from private plan to private plan or to a government plan, and no single source houses information on all patient encounters. United States nongovernmental databases may obtain their data through third-party insurer claims.

Published by Elsevier Inc.
ISSN 0161-6420/14/$ - see front matter
http://dx.doi.org/10.1016/j.ophtha.2013.11.038

Table 1. Checklist for the Reporting of Studies That Use Claims Data*

Description of data source	Was a description of the sociodemographic characteristics and health care profile of the population provided?
	Did the investigators acknowledge limitations of services provided due to type of insurance or plan type of the enrollees, benefit design, and how providers are reimbursed for services studied?
Checks of data quality	Sources of unreliable data include changes in reporting/coding practices over time and in reporting resulting from changes in reimbursement, and services may be inadequately captured if not covered by the plan.
	Did the investigators explain how they handled missing and out-of-range values?
	Did the investigators explain how they handled duplicate claims and inconsistencies (differences in age of same patient on different claims)?
	Did the investigators compare the reported rates of disease or use with established norms or other data sources?
	If other researchers have studied the reliability and validity of the data source used, those should be cited.
	Have the necessary linkages among data sources or sites of care been carried out appropriately?
	Is there an explanation of how member eligibility was determined?
Sample selection	Is there a sample selection figure to easily show readers the numbers of enrollees included and excluded and for what reasons?
	Is there is justification provided for using the chosen inclusion/exclusion criteria for selecting beneficiaries for the study sample?
	Is there a transparent listing of all of the ICD-9-CM and CPT codes used in the study?
	Were enrollees who were noncontinuously enrolled in the health plan during the entire study period included in the analysis?
Analysis	Is the data analysis plan clearly described?
	Were research hypotheses generated a priori or were the findings generated the result of unsystematic data exploration?
	Did the investigator provide a cogent rationale for the study design chosen, in light of the data, setting, and research questions? Are limitations of the study design chosen clearly delineated to the reader? Examples of potential biases include selection bias, maturation, and regression to the mean.
	For studies reporting treatment effects, was there a control group created to compare against the group receiving the intervention?
	Did the investigators censor subjects, and, if so, did they explain how this may affect the sample selection or generalizability of the cohort?
	Are the end points or outcomes clearly defined on the basis of diagnosis or procedure codes or other criteria?
	Did the investigators justify the definition of the end points they chose for the analysis or cite other sources who used similar criteria?
	Were sensitivity analyses performed to explore the impact of changing the criteria for study inclusion or the definition of the outcome(s) of interest?
	Is there a temporal relationship between the exposure and the outcome of interest (did the researchers require the exposure to come before the outcome)?
Statistics	Were important confounding factors identified and adjusted for in the analyses either by stratification of the sample by the confounding variable or by the use of multivariable statistical techniques?
	What sort of risk adjustment was performed? Did the investigators account for differences in sociodemographic characteristics, medical comorbidities, disease severity?
	Were adequate tests of the statistical assumptions performed? Examples include testing for multicollinearity and adjustment for multiple comparisons.
Discussion/ conclusions	Did the investigators provide a rationale for the study findings in light of the existing literature? Were alternative explanations for the findings offered?
	Did the investigators comment on the clinical or economic relevance of the study findings because statistical significance may not necessarily translate into clinical significance?
	Did the authors address concerns about the generalizability of study findings to other groups?
Funding sources	Were the funding sources for the analyses clearly identified? Did the funding sources participate in designing or conducting the study?

CPT = Current Procedural Terminology; ICD-9-CM = International Classification of Diseases, 9th Revision, Clinical Modification.
*Adapted from Motheral B, Brooks J, Clark MA, et al. A checklist for retrospective database studies—report of the ISPOR Task Force on Retrospective Databases. Value Health 2003;6:90–7.[2]

The types of information contained in claims databases can vary considerably. Nearly all such databases record patients' diagnosed medical conditions by using International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM),[3] billing codes and diagnostic and therapeutic procedures according to Current Procedural Terminology 4 (CPT-4)[4] billing codes. Commonly documented demographic variables include age at enrollment in the plan, sex, race, ethnicity, education, and income. Some databases are linked to outpatient pharmacy records and results of outpatient laboratory testing; some capture the types of providers delivering the care, the site of care delivery (e.g., inpatient, outpatient, ambulatory surgical center), and the costs or charges of each service. Each data source typically contains a dictionary with a complete listing of variables.

Claims databases that have been used to study patients with ophthalmologic conditions include a 5% national sample of Medicare beneficiaries (Centers for Medicare and Medicaid Services); the Veterans Health Administration system National Patient Care Database, the Clinformatics Data Mart database (OptumInsight, Eden Prairie, MN); the PharMetrics Patient-Centric database (PharMetrics, Inc., Watertown, MA); and the Market Scan Commercial Claims and Encounter Data (Truven Health Analytics, Ann Arbor, MI). Other datasets include Medicare claims data linked with clinical, pharmacy, or survey data. Examples include the National Long-Term Care Survey, Medicare Current Beneficiary Survey, and Medical Expenditure Panel Survey. In addition, some non–U.S. administrative datasets capture care of patients in other countries.