How Healthcare Studies Use Claims Data

Bryan Burton and Paul Jesilow[*]

Department of Criminology, Law and Society, School of Social Ecology, University of California, Irvine, CA 92697-7080, USA

Abstract: Claims data have become common during the past two decades. The electronic records include information entered on bills (claims) submitted by healthcare providers to third-party payers. They are an attractive data source; however, they contain limitations that threaten the validity of studies that use them. We reviewed 168 studies that employed claims data, published during 2000-2005 in five healthcare journals, to investigate how claims data are being used and whether their use is appropriate. Healthcare studies in our sample used claims data to select a sample, to establish healthcare costs, to determine whether specific treatments or procedures had been provided, to ascertain the cost-effectiveness of services, and to establish their accuracy as a stand-in for other measures. Most studies appropriately used claims data; however, there was a sizable percentage that used claims data in an inappropriate or questionable fashion.

Keywords: Electronic data, claims data, administrative data, measures, upcoding.

INTRODUCTION

Hundreds of published healthcare studies during the past 25 years have employed claims data. Their use in research grew during the 1990s as the data became more common. Claims data are electronic records of transactions that provide information on healthcare encounters that supposedly have taken place between providers and patients. These data are information entered on bills (claims) and submitted by hospitals, physicians, home healthcare agencies, durable medical equipment providers, independent testing facilities, clinics, pharmacies and other medical professionals that are submitted to public (e.g. Medicare and Medicaid)[1] and private insurance entities (e.g. Anthem Blue Cross and Blue Shield). These electronic data sets, also known as billing and occasionally as administrative data, have been used to investigate a broad array of medical issues, from the use of comorbidity indices to predict risk of death [1] to the use of children's mental health services [2].

Claims data have been primarily used in studies that focus on access to healthcare and quality issues [3]. Examples of access studies include the effects of different societal or policy changes on patient use of health services. One study, for example, used claims data to examine the effect of a reduction of welfare payments on mental health service use [4], while another study explored Medicare beneficiaries use of prescription drug discount cards [5]. Examples of quality of care studies include the use of claims data to establish benchmark left ventricular ejection fraction test rates for Medicare recipients [6], while another study assessed the association between office systems and variations in diabetes care [7].

Claims data are alluring to researchers because they offer numerous advantages. They are anonymous, abundant, economical, and widely available in electronic format [8]. This has made their use widespread in research studies, particularly as a replacement for medical records. They may provide useful information; however, they have limitations that may weaken the validity of some studies that utilize them and undermine healthcare policy based on such research [3]. Their use in research studies as a result is not always appropriate.

In this piece, we present the results of a review of research articles that utilized claims data. We were interested in determining how claims data are used in healthcare studies and under what circumstances such use is appropriate.

METHODOLOGY

We initially examined 1,956 research articles that were published during a six-year span (2000-2005) in five healthcare journals (*The New England Journal of Medicine, American Journal of Medical Quality, Medical Care, Medical Care Research and Review*, and *Health Care Financing Review*) to estimate the extent of use of claims data in healthcare studies. The selection process of the journals and articles, as well as our review process, are described in greater detail elsewhere [3]. We concluded that the extent of use of the electronic data in the five journals had stabilized, although there were great differences between the percentages of each journal's articles that used claims data [3].

Our initial analysis reported 143 studies that utilized claims data, but we had excluded studies that employed other administrative data, such as hospital discharge abstracts, registries and a combination of other measures [3]. The present inquiry includes such studies if we were able to determine that the administrative data included measures

*Address correspondence to this author at the Department of Criminology, Law and Society, School of Social Ecology, University of California, Irvine, CA 92697-7080, USA; Tel: 949-856-0677; Fax: 949-824-3001; E-mail: paul.jesilow@uci.edu

[1] Medicare is a U.S. tax-supported health insurance program for the elderly and disabled. Medicaid is a state-operated health insurance program for the poor; it is jointly funded by the federal government and the states.

derived from claims. This increased the number of studies that utilized claims in our sample to 168 (of the 1,956). Most of these studies used U.S. data (164 of the studies used data derived from U.S. data bases; three of the studies used data from Canadian sources; and, one study used data from Taiwan). We reviewed these 168 studies to determine how the data were used and whether that use was appropriate.

RESULTS

The vast majority of the studies examined in the current research utilized claims data to establish a sample; however, most of these studies used claims data in some additional fashion. Only 14 of the 168 studies (8.3%) used claims data solely to select a sample. Claims databases are an appropriate and useful means for finding sizable groups of patients, although they do have a limitation. Diagnoses are not always correctly recorded in claims data. "Upcoding," for example, leads to errors in claims data. Billing clerks may indicate in claims that the patients had more expensive diagnoses or were provided more expensive procedures or products than the ones actually provided. Patients, as a simple illustration, receive generic drugs, but the claims indicate that name brands were provided [9].

Researchers, who used claims data solely to select a sample of individuals who had a specific diagnosis, confirmed the diagnosis by checking the medical record in order to eliminate errors found in claims data. These researchers were usually attempting to locate sizable groups of patients with rare conditions, who might be difficult to locate by other means [3]. Once the sample was located, it was a much easier task for the researchers to validate the diagnosis than the effort it would have required to locate the sample by other means. Elaine Hylek and her colleagues [10], for example, wanted to locate a sample of patients who had ischemic strokes. They electronically searched claims databases to initially select their sample of individuals. They then reviewed the individuals' medical records to validate the diagnosis. Without such efforts, the use of claims to select a sample may not always be appropriate. We discuss this issue (below) with respect to the use of claims data as an appropriate measure to determine utilization or effectiveness of interventions.

Forty-five studies (26.8%) used claims data to establish estimates of the costs for various healthcare items. The purpose of a claim is to collect payment, so it is convenient for researchers to consult fee schedules and reimbursement data to establish cost estimates [e.g., see 11]. Coding errors and questionable billing practices may contaminate the costs of a specific diagnosis, but the financial outlays for the diagnosis are nevertheless real. Upcoding is illustrative; such billing activities are illegal, but the resultant increased costs for the recorded diagnosis of interest are nonetheless real [9].

Eighty-four studies (50%) employed claims data as measures of healthcare items, commonly to determine whether specific treatments or procedures had been provided to patients. Generally, the researchers assume that a bill (claim) for the healthcare activity indicates it was provided; the absence of a claim indicates to them that it was not. Eleven of these 84 studies also used claims data to ascertain the cost-effectiveness of the procedures and/or services. These 11 studies used claims to establish which procedures

had been delivered, but also to establish the cost of the various services and procedures. Michael Sokol and his colleagues [12], for example, evaluated the impact of adherence to medication on healthcare utilization and costs for four chronic conditions (diabetes, hypertension, hypercholesterolemia and congestive heart failure). Pharmacy claims data and medical claims data were used to measure drug cost, medical cost, and utilization. Individuals who had their prescriptions refilled were assumed to be adhering to the drug utilization under the direction of their physicians. Such use of claims data is not always appropriate, a matter which we discuss in the next paragraphs.

Claims data can be an appropriate measure to determine utilization or effectiveness (including cost-effectiveness) of interventions if prior research has established that the claims data are a good stand-in for medical records. Indeed, some of the studies in our sample utilized claims data in order to determine their validity as measures of various conditions. Twenty-five studies (14.9%) in our sample, for example, compared claims data with some other data source, primarily medical records, but also registries. These "concordance" studies were designed to determine the validity of claims data as a stand-in for the other measures. The researchers were generally interested in establishing the likelihood of a diagnosis or procedure recorded in the claims data also appearing in the medical record (known as "sensitivity"). But, they were also interested in the likelihood that a diagnosis or procedure not present in the claims data would also not be recorded in the medical record (known as "specificity") [13]. We consulted concordance studies (including ones not in our sample) to determine whether the use of the claims data was appropriate for each of the 84 studies in our sample that used claims to determine utilization or effectiveness (73 studies) or cost effectiveness (11 studies).

We classified as "likely appropriate" 54.8% (46 of 84) of the studies that utilized claims as stand-ins for other measures. These studies primarily used claims as measures of major procedures and conditions; our review of concordance studies indicated that claims do a relatively good job as measures of these items. For example, Medicare claims data have been compared with surgeries recorded by the Surveillance, Epidemiology and End Results[2] registries. The two data sources were in agreement more than 95% of the time for mastectomies and 91% of the cases for breast-conserving surgery [14]. Another illustrative study compared data for surgical procedures abstracted from the medical records of women diagnosed with node-negative breast cancer with both hospital discharge abstracts and physician claims and reached similar results; the two databases accurately reproduced the information recorded in the patients' charts [15]. Other procedures that have high rates of agreement between medical charts and administrative data

---

[2]The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute works to provide information on cancer statistics in the U.S. Population. The SEER currently collects and publishes cancer incidence and survival data from population-based cancer registries, which account for approximately 28% of the U.S. population.

include appendectomies (94.3%; ICD-9-CM[3] Codes 470, 471) and biopsies of bone marrow (87.5%; ICD-9-CM Code 4131) [16].

Based on our review, we classified as "likely inappropriate" 38% (32 of 84) of the studies that utilized claims as stand-ins for other measures. These were studies that used claims data as measures for health conditions that may result in conflicting diagnosis or as measures for minor procedures. The rate of agreement between claims data and medical records, according to the concordance studies we reviewed, is poor for conditions that may result in conflicting diagnoses, such as alcohol or drug abuse. One healthcare provider may consider as abuse, for example, ingestion of a specific amount of liquor each day, while another provider may find the same situation acceptable. Such disagreement, it is believed, impacts how such conditions are reported in medical records and claims data; for example, the likelihood in one study of the diagnosis alcohol/drug abuse (ICD-9-CM Codes 291-305.93) being recorded in the claims data and also appearing in the ambulatory medical records was only 20% [17]. Minor procedures, such as some injections or local excisions, also reveal low levels of agreement between medical records and hospital administrative data (upon which claims data are based); only 7.3% for insertion of nasogastric tube (ICD-9-CM Codes 9607, 9606) and zero agreement for insertions of indwelling urinary catheters (ICD-9-CM Code 5794) [16]. Using claims data as measures of health conditions that may result in conflicting diagnosis or as measures for minor procedures is inappropriate.

We classified as "questionable" 7.1% (6 of 84) of the studies that utilized claims as stand-ins for other measures. Each of these six studies used claims data as measures for multiple medical conditions and circumstances. Michael Shwartz and his colleagues, for example, examined Medicare data covering both inpatient admissions and outpatient visits for 15 medical conditions to estimate hospitalization rates, disease-based hospitalization rates and disease prevalence [18]. Most of the medical conditions of interest to Shwartz and his colleagues, according to the concordance studies we reviewed, have high rates of agreement between the claims data and the medical records (such as about a 85% sensitivity rate for diabetes, ICD-9-CM codes 250-250.99 [13]). But, other medical conditions of interest to the researchers have low rates of agreement (such as a 58% sensitivity rate for peripheral vascular disease, ICD-9-CM codes 440-442.9, 433.9 [13]). We labeled such studies as "questionable" because the claims data were not uniformly appropriate or inappropriate.

CONCLUSION AND DISCUSSION

Most studies appropriately used claims data. We categorized one hundred thirty studies (77.4%) as using claims data in a "likely appropriate" manner. Studies that used claims data to solely select a sample or to establish costs for a specific diagnosis were generally on solid ground. Claims data were also used appropriately by researchers who tested the validity of the claims as measures by comparing them with other data sources, most commonly medical records (from which claims data are normally derived).

Researchers were often on less solid ground when they used claims data to determine the utilization or effectiveness (including cost-effectiveness) of interventions. We did categorize more than half of such studies as using the data appropriately; concordance studies indicated that the claims were valid measures of the items of interest. But we also categorized nearly half of such studies as "inappropriately" or "questionably" using claims data; the concordance studies indicated that claims were not uniformly good measures of the items of interest. It is important to reiterate that we are not judging the studies, but we are evaluating whether the use of claims data were appropriate.

Researchers and policymakers need to be hesitant about accepting the results of studies that use claims data that are not valid measures. Researchers contemplating the use of claims data need to be aware of the accuracy of them as a stand-in for other measures. Researchers and policymakers would benefit from a single source that would provide them with established concordance rates (sensitivity and specificity) for as many items as possible. We have set this as our next task.

REFERENCES

[1]   Yan Y, Birman-Deych E, Radford MJ, Nilasena DS, Gage BF. Comorbidity indices to predict mortality from Medicare data: results from the national registry of atrial fibrillation. Med Care 2005; 43(11): 1073-7.

[2]   Larson MJ, Miller K, Sharma S, Manderscheid R. Children's mental health services in fee-for-service Medicaid. Health Care Financ R 2004; 26(1): 5-22.

[3]   Ferver K, Burton B, Jesilow P. The use of claims data in healthcare research. Open Public Health J 2009; 2: 11-24.

[4]   Steele LS, Glazier RH, Lin E, Austin PC, Mustard CA. Measuring the effect of a large reduction in welfare payments on mental health service use in welfare-dependent neighborhoods. Med Care 2005; 43(9): 885-91.

[5]   Eppig FJ, Poisal JA. Medicare beneficiary's use of prescription drug discount cards, CY 2002. Health Care Financ R 2003; 25(2): 91-4.

[6]   Wu B, Pope GC. Left ventricular ejection fraction test rates for Medicare beneficiaries with heart failure. Am J Med Qual 2002; 17(2): 61-6.

[7]   Ellerbeck EF, Engelman KK, Williams NJ, Nazir N, Markello SJ. Variations in diabetes care and the influence of office systems. Am J Med Qual 2004; 19(1): 12-8.

[8]   Hicks J. The potential of claims data to support the measurement of healthcare quality. Santa Monica, CA: Rand Corporation 2003.

[9]   Jesilow P. The effects of fraud on the evaluation of health care. Health Care Anal 2005; 13(3): 239-45.

[10]   Hylek EM, Go AS, Chang Y, *et al.* Effect of intensity of oral anticoagulation and stroke severity and mortality in atrial fibrillation. N Engl J Med 2003; 349(11): 1019-26.

[11]   Morris AM, Flowers CR, Morris KT, Schmidt WA, Pommier R, Vetto JT. Comparing the cost-effectiveness of the triple test score to traditional methods for evaluating palpable breast masses. Med Care 2003; 41(8): 962-71.

[12]   Sokol MC, McGuigan KA, Verbrugge RR, Epstein RS. Impact of medication adherence on hospitalization risk and healthcare cost. Med Care 2005; 43(6): 521-30.

[13]   Fisher ES, Whaley FS, Krushat WM, *et al.* The accuracy of Medicare's hospital claims data: Progress has been made, but problems remain. Am J Public Health 1992; 82(2): 243-8.

[14]   Du X, Freeman Jl, Warren JL, Nattinger AB, Zhang D, Goodwin JS. Accuracy and completeness of Medicare claims data for surgical treatment of breast cancer. Med Care 2000; 38(7): 719-27.

---

[3]The ICD-9-CM indicates the International Classification of Diseases, 9th Revision, Clinical Modification.

[15]   Pinfold PS, Goel V, Sawka C. Quality of hospital discharge and physician data for type of breast cancer surgery. Med Care 2000; 38(1): 99-107.

[16]   Quan H, Parsons GA, Ghali WA. Validity of procedure codes in international classification of diseases, 9th revision, and clinical modification administrative data. Med Care 2004; 42(8): 801-9.

[17]   Fowles JB, Fowler EJ, Craft C. Validation of claims diagnoses and self-reported conditions compared with medical records for selected chronic diseases. J Ambul Care Manage 1998; 21(1): 24-34.

[18]   Shwartz M, Pekoz EA, Ash AS, Posner MA, Restuccia JD, Iezzoni LI. Do variations in disease prevalence limit the usefulness of population-based hospitalization rates for studying variations in hospital admissions? Med Care 2005; 43(1): 4-11.