Bias due to False-Positive Diagnoses in an Automated Health Insurance Claims Database

Stephan F. Lanes[1] and *Cynthia de Luise*[2]

1   Boehringer Ingelheim Pharmaceuticals Inc., Ridgefield, Connecticut, USA
2   Pfizer Inc., New York, New York, USA

Abstract

Background and objective: Automated database studies have become a cornerstone of drug safety assessment. To assess the reliability of automated data, we compared the hospitalisation and mortality rates among three similar studies of automated healthcare databases in North America.

Methods: Similar protocols were used to identify patients diagnosed with chronic obstructive pulmonary disease (COPD) who were treated with inhaled bronchodilators or inhaled corticosteroids in the Saskatchewan Health Database (SHD), the Kaiser Permanente Medical Care Program (KPMCP) of Northern California, and a proprietary automated insurance claims database available from i3® (formerly Ingenix®). Automated data were used to compute incidence rates of total hospitalisation, cardiovascular (CV) hospitalisation and hospitalisation due to several specific types of CV outcomes. Record linkage with registries of vital statistics was used to identify deaths, obtain death certificates, and compute rates of total mortality, CV mortality and deaths due to certain CV outcomes. We compared rates in the i3® population with rates in the other two populations using age-adjusted rate ratio estimates and 95% CIs.

Results: The i3® cohort had approximately one-half the rates of total mortality, CV mortality and total hospitalisations, but twice the rate of CV hospitalisations, compared with each of the other two database cohorts.

Discussion: The unexpectedly higher rates of CV hospitalisations in the i3® population are inconsistent with its lower CV mortality, total mortality and total hospitalisation rates. This discrepancy is not readily explained by a higher prevalence of CV disease or procedures, random variation or confounding. Instead, high CV hospitalisation rates in the i3® population are consistent with a high rate of false-positive diagnoses recorded on insurance billing claims.

Conclusion: These results underscore the importance of ensuring valid endpoints in automated claims databases.

Background

In the wake of mounting concerns about the adequacy of drug safety monitoring, regulators and pharmaceutical companies are looking for better strategies to identify adverse effects in a more timely manner.[1] One approach is to increase the use of automated healthcare databases to conduct epidemiological safety surveillance for new medications.[2] Automated healthcare databases typically contain administrative records of medications dispensed, as well as physician visits and hospitalisations. Such data can be analysed quickly to obtain statistics relating drugs and health outcomes.

However, administrative data are not usually collected for research purposes. Furthermore, because researchers may not be involved with the design, collection or automation of the source data, errors may exist in the automated data that are difficult to recognise. This article reports the findings of a study in three automated healthcare databases.

Methods

We assessed the rates of hospitalisations and mortality due to certain types of cardiovascular (CV) illness in patients with chronic obstructive pulmonary disease (COPD). To enhance reliability, we used a similar protocol in three popular automated healthcare databases in North America: the Saskatchewan Health Database (SHD), the Kaiser Permanente Medical Care Program (KPMCP) of Northern California, and a proprietary database comprising insurance claims from UnitedHealth Group® and available from the consulting group i3® (formerly Ingenix®).[3-5]

The SHD is developed and maintained by Saskatchewan Health, a provincial government department.[3] Universal health insurance is available to all of the approximately 1 million residents of Saskatchewan, Canada and eligibility is not dependent on socioeconomic status. The KPMCP in Northern

California has about 3 million members and is owned by Kaiser Permanente, which operates the largest group-model health maintenance organisation (HMO) in the US.[4] KPMCP features a broadly representative population, owing to the prevalent membership in its geographic coverage area, which includes many individuals aged >65 years. The i3® database is created from UnitedHealth Group® diversified insurance plans throughout the US, which are mainly independent practice associations.[5] The population includes about 4 million members, is primarily an employed population, and is less stable in terms of duration of membership than the other two databases.[5]

Each study identified populations with COPD as individuals aged ≥40 years who had inpatient or outpatient records containing diagnosis codes for chronic bronchitis (International Classification of Diseases, 9th Edition [ICD-9] code 491), emphysema (492), or chronic airway obstruction (496), and who also had received at least two prescriptions for respiratory medications (inhaled bronchodilators or inhaled corticosteroids). Hospitalisations were identified by records indicating an inpatient hospital stay. Endpoints in each database included total hospitalisations (for any reason) and hospitalisations containing a primary diagnosis code indicating any CV diagnosis, and specific diagnoses of myocardial infarction (MI), congestive heart failure (CHF), arrhythmia, stroke or pulmonary embolism.

Mortality was identified through linkage with external vital statistics registries. CV deaths were identified by the underlying cause of death coded on the death certificate. In two databases (KPMCP, SHD), we validated the hospitalisation diagnoses in automated data using a sample of patient records. In one database (i3®), because of the relatively high cost of obtaining charts, validation was confined to arrhythmia diagnoses. Covariates included age, sex, concomitant medications and concomitant morbidities. In addition, each study population was sampled

to determine smoking history and body mass index. Data vendors undertook ethical review to ensure patient confidentiality and compliance with all applicable regulations.

The calendar period of study was similar within each database: 1 January 1997–12 December 2000 for SHD, 1 January 1996–12 December 1999 for KPMCP, and 1 January 1997–30 June 2001 for i3®. Person-time at risk started with completion of cohort inclusion criteria within the study period, and subjects were followed until the enrolment end date, date of death or study end date, whichever came first. We computed incidence rates as the number of events divided by the person-time at risk. Details of these analyses are described elsewhere.[6-9]

To enable valid comparisons between databases, we obtained from each vendor tabulations of aggregate data for each endpoint, stratified by age. We reviewed age-specific hospitalisation and mortality rates in the three populations, and we compared the i3® population directly with each of the other two populations, by computing Mantel-Haenszel rate ratio estimates adjusted for age (in 10-year categories). We assessed the precision of the estimates by the width of the 95% CIs.

Results

The COPD populations comprised 11 493 individuals in the SHD; 45 966 in the KPMCP; and 18 894 in i3® (table I). Mean duration of follow-up was 2.7 years for SHD, 2.8 years for KPMCP, and 1.3 years for i3®. Sex distribution was similar and balanced across the three cohorts, with the proportion of males ranging from 49% to 55%. Age distributions revealed important differences between the databases. The SHD includes almost all the residents in the province and has the oldest population, with 74% of the COPD population aged ≥65 years. In the KPMCP, most COPD patients (53%) are also aged ≥65 years, while i3® has the youngest population, with only 15% of the COPD population aged ≥65 years. Consistent with the younger age of those in the i3® cohort, use of CV medications was less prevalent than in the KPMCP and SHD cohorts.

Age-adjusted rate ratio estimates comparing i3® with the other populations are presented in figure 1. The i3® cohort demonstrated about half the mortality rate compared with the SHD and the KPMCP cohorts. Lower rates in i3® were observed for total mortality, total CV mortality and for each of the CV causes of death individually. As with mortality, the rate of total hospitalisation in i3® was substantially lower than total hospitalisation rates in the KPMCP or the SHD. However, for CV hospitalisations the pattern was unexpectedly reversed. While the rates of total mortality, CV mortality and total hospitalisations were about 50% *lower* in i3®, the CV hospitalisation rate was approximately *twice* CV hospitalisation rates in the KPMCP or the SHD. Rates of hospitalisation due to each specific type of CV outcome examined, including MI, CHF (figure 1), as well as arrhythmia, stroke and pulmonary embolism (data not shown), were also elevated approximately 2-fold in i3® compared with KPMCP and SHD. Figure 1 shows values of effect estimates for CV hospitalisations that would have been expected if these effect estimates had been similar in magnitude to the effect estimates for total hospitalisation and mortality. With regard to random error, the studies are large and most of the effect estimates

Table I. Age distribution in chronic obstructive pulmonary disease cohorts

Age group (years)	i3®a [n (%)]	KPMCP [n (%)]	SHD [n (%)]
40–49	3 625 (19)	6 757 (15)	547 (5)
50–59	8 332 (44)	9 285 (20)	1 271 (11)
60–64	4 104 (22)	5 700 (12)	1 127 (10)
65+	2 833 (15)	24 224 (53)	8 548 (74)
Total	**18 894 (100)**	**45 966 (100)**	**11 493 (100)**

a   i3® is a proprietary automated insurance claims database (formerly Ingenix®).

KPMCP = Kaiser Permanente Medical Care Program; **SHD** = Saskatchewan Health Database.
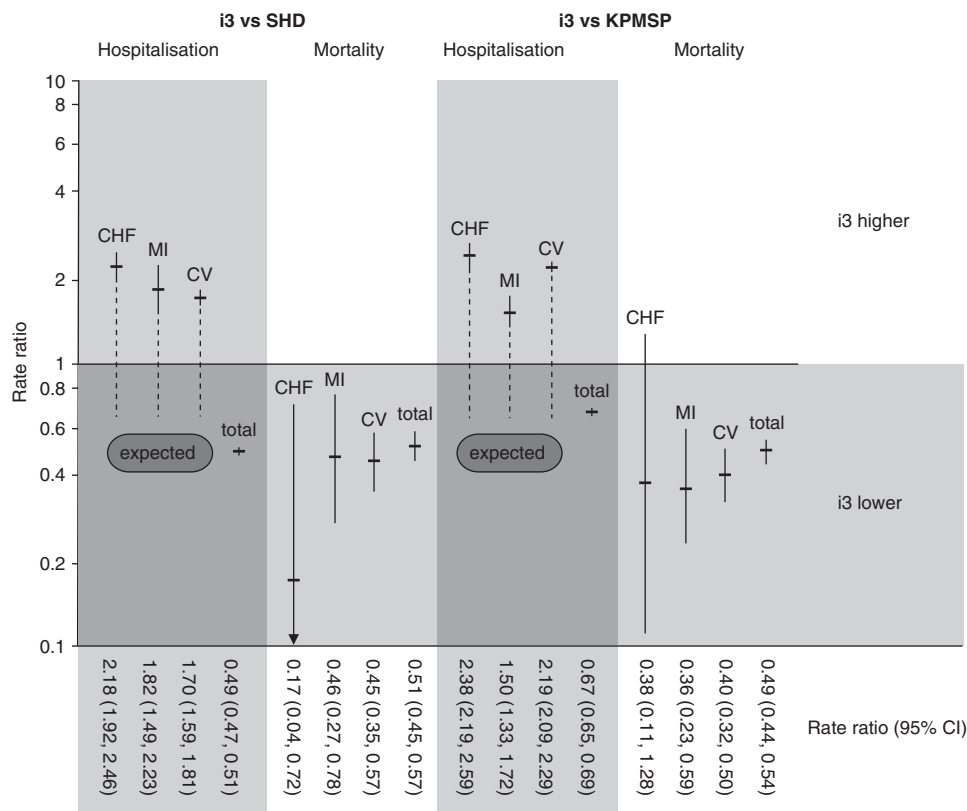
Fig. 1. Incidence rate ratios of mortality and hospitalisations in the i3® (formerly Ingenix®) automated insurance claims database compared with those in the Kaiser Permanente Medical Care Program (KPMCP) of Northern California and the Saskatchewan Health Database (SHD). **CHF** = congestive heart failure; **CV** = cardiovascular; **MI** = myocardial infarction.

have good precision, particularly for hospitalisation endpoints.

Validation studies based on a sample of cases in KPMCP and SHD showed good agreement between automated data and hospital records, especially for MI and CHF, each of which had >95% agreement. Validation was attempted by i3®, but only for a subset of arrhythmia cases for which there were both diagnostic codes and new treatment codes (no related claims in the baseline period), and whose profiles were reviewed and judged to be probable arrhythmias. Medical records were obtained for 62% of these individuals. One would expect better accuracy for these subjects than for subjects identified only by a diagnosis code, as used in the study. Nevertheless,

in this subgroup, 18% were prevalent cases, and the type of arrhythmia listed in the automated data was confirmed in 52% of the records.

Discussion

Application of similar epidemiological protocols to the i3®, KPMCP and SHD databases offers an unusual opportunity to compare results between three widely used North American automated healthcare databases. Substantially lower rates of mortality and hospitalisations in the i3® population are consistent with its demographic profile and suggest that it is a healthier population, in part possibly because of a manifestation of the healthy-worker

effect, i.e. selection factors in the i3® insurance claims database related to employment.

Unexpectedly, the i3® population shows markedly higher rates of CV hospitalisations than each of the other two populations. These elevated rates of CV hospitalisations are anomalous with the demographics of the i3® population, and also internally inconsistent with its lower total hospitalisation rate and low total mortality and CV mortality rates. The increased rate of CV hospitalisations could be explained if the i3® population were at increased risk of CV disease or had more elective CV procedures. However, if the i3® population were at increased risk of CV disease, we would expect it to also have higher rates of CV mortality. Instead, the i3® population has lower rates of CV mortality than the other populations. If excess CV hospitalisations were due to elective CV procedures, these hospitalisations would be reflected in the overall hospitalisation rate. More generally, since CV disease is the most common reason for hospitalisation in the US, and is even more common in COPD, a markedly elevated rate of CV hospitalisations would also mean an elevated rate of total hospitalisations.[10] Instead, the i3® population has the lowest hospitalisation rate. Therefore, the higher rates of CV hospitalisation in i3® do not appear to be because this population is at increased risk of CV disease or has more frequent hospitalisations for elective procedures.

We considered that the effect estimates for CV hospitalisations might be inaccurate. When random error is evaluated, the hospitalisation rate ratio estimates are rather precise and the pattern of results is not easily explained by random statistical variability. With regard to systematic error, the i3® cohort was younger than the other cohorts, but age was controlled in the analysis. Moreover, because i3® comprised the youngest population, any residual confounding by age would bias the CV hospitalisa-

tion effect estimates downwards and could not explain the higher rates observed in this cohort.

Another possible source of bias is errors in identifying endpoints or 'outcome misclassification'. Diagnoses recorded in automated hospitalisation records were validated at KPMCP and SHD, and showed good reliability of ≥95% for key endpoints. The i3® validation study did not sample cases by diagnosis code, but instead was restricted to probable cases defined by combinations of diagnosis and procedure codes, so the results do not enable calculation of error rates associated solely with diagnosis codes in the automated data. Still, only about one-half of the diagnoses in the probable cases in the automated data could be confirmed, consistent with a higher false-positive error rate in the i3® CV hospitalisation diagnoses.

False-positive diagnosis errors in the automated database used by i3® are confirmed by two previous validation studies.[11,12] One study found that "the submission of a diagnosis of hypertension on a single claim form is not a valid indicator of the presence of hypertension".[11] The other study concluded that "in itself the claims data was inadequate for case verification".[12] Recently, others compared the i3® automated claims data with medical records and concluded that "identification of disease states by diagnostic or procedural codes alone is likely to produce great inaccuracy".[13] Therefore, validation studies support the idea that a CV diagnosis code in the i3® automated hospital claims data often does not indicate a CV hospitalisation.

In considering the source information for the endpoints studied, errors in hospitalisation diagnoses would not affect rates of total hospitalisations or mortality. The reason is that total hospitalisation rates are computed on the basis of the existence of a hospital claim and without regard to information contained in the diagnosis fields. Similarly, mortality rates rely on data from death certificates, which also would not be affected by diagnosis coding

errors on billing claims. Therefore, false-positive errors in hospital diagnoses in i3® automated data would spuriously increase CV hospitalisation rates in the i3® study, but would not affect total rates of hospitalisation, total mortality, or CV mortality.

Initially, we had little reason to suspect that hospital diagnosis coding errors would occur more frequently in the i3® database. Further evaluation revealed that error rates in hospital diagnosis could differ in the i3® database because diagnoses in this database represent different source data from those in the other databases. Whereas SHD and KPMCP databases contain the primary hospital *discharge* diagnosis,[8,9] the i3® database contains the principal *billing* diagnosis recorded on an insurance claim form (UB-92).[5] Because the KPMCP is a group-model HMO and SHD contains data obtained under a national health plan, these administrative data do not represent billing claims. Possible reasons for inaccurate coding of diagnoses on i3® insurance claim records include coding to a diagnosis that had not been established or a diagnosis that was being ruled out, and miscoding for administrative reasons related specifically to billing issues.[11-13] One coding problem associated with billing claims is referred to as 'upcoding', which describes medical miscoding related to increased financial reimbursement.[14] Upcoding can vary by diagnosis, type of hospital, insurer and healthcare system.[14,15] In discussing accuracy of hospitalisation diagnoses in the UnitedHealth Group® claims data used by i3®, a recent text points out: "in a discounted fee-for-service plan, there may be a financial incentive to code diagnoses according to the most profitable reimbursement schedules".[16] The endpoints in this study were CV illnesses that would be associated with substantial costs. Therefore, there is a basis for postulating that financial incentives or other billing-related issues could be responsible for more frequent diagnosis errors in the i3® database.

False-positive diagnosis errors in i3® would explain the high rate of CV hospitalisations in the i3® cohort, and such errors are corroborated in several validation studies. This theory would suggest that other important diagnoses could also be over-represented in the i3® database, and that there may be a complementary under-representation of less expensive diagnoses. Such analyses would be useful to explore but are beyond the scope of this study, which was confined to CV diagnoses.

If elevated CV hospitalisation rates in this i3® study were due to false-positive coding errors, the proposed mechanism implies that such errors would probably not be confined to this study alone or to the i3® database. Diagnosis errors could be prevalent in insurance billing claims. Although the problem of inaccurate diagnosis codes is recognised, it is not widely appreciated and this study provides a new perspective on the extent of bias that can be introduced by this source of error. This matter is of growing importance because of the increasing use of the i3® database and other automated data to monitor the safety of new drugs by pharmaceutical companies and the US FDA.[17] In studies relying on automated data to compare cause-specific hospitalisation rates among people exposed and unexposed to a particular drug, an important drug safety hazard could be overlooked because false-positive diagnosis errors, even if they occur at the same rate in the groups being compared (i.e. nondifferential outcome misclassification), will tend to bias rate ratio estimates towards the null value (RR = 1.0).[18] If automated diagnoses were used to control for comorbidities, the impact would be a loss in control of confounding, which could bias effect estimates in either direction.[18] Ultimately, the direction and magnitude of bias arising from misclassification of diagnoses are not straightforward and would vary depending on the endpoint, the groups being compared, method of case ascertainment and other aspects of a particular study.

Conclusion

Diagnoses in automated insurance claims data and the i3® database in particular may be inaccurate to an appreciable extent. One solution is to validate automated claims data using medical records. We would advise researchers who use automated databases to consider carefully the extent to which a particular database and endpoint has been validated.[16] The cost and time required to validate automated records and patient privacy considerations will vary and can be barriers to validation. Indeed, the need to comprehensively validate large numbers of potential cases could undermine the very efficiency that is the chief reason for conducting automated claims-based studies. Nevertheless, the cost of not validating diagnoses could be results that are unreliable as a basis for decision making.

Acknowledgements

References

1. Harris G. Drug safety system is broken, a top FDA official says. New York Times 2005 Jun 9; Late Edition-Final, Sect. A: 24 (col 4)
2. Fuhrmans V. Early-warning tool for unsafe drugs. Wall Street Journal 2005 Apr 28, D4
3. Downey W, Stang MR, Beck P, et al. Health services databases in Saskatchewan. In: Strom BL, editor. Pharmacoepidemiology. 4th ed. New York (NY): John Wiley & Sons Ltd, 2005: 295-310
4. Selby JV, Smith DH, Johnson ES, et al. Kaiser permanente medical care program. In: Strom BL, editor. Pharmacoepidemiology. 4th ed. New York (NY): John Wiley & Sons Ltd, 2005: 241-60
5. Shatin D, Rawson NSB, Stergachis A. UnitedHealth Group. In: Strom BL, editor. Pharmacoepidemiology. 4th ed. New York (NY): John Wiley & Sons Ltd, 2005: 271-80
6. Lanza L, Norman H, Quinn SG, et al. A cohort study of cardiovascular conditions in COPD. Auburndale (MA): Ingenix Epidemiology, a UnitedHealth Group Company, 2003 Jul 29 (data on file)
7. de Luise C, Lanes S, Curkendall S, et al. Cardiovascular morbidity and mortality among persons with chronic obstructive pulmonary disease [abstract]. Pharmacoepidemiol Drug Saf 2004; 13: S238
8. Sidney S, Sorel M, Quesenberry CP, et al. COPD and incident cardiovascular disease hospitalizations and mortality: Kaiser Permanente Medical Care Program. Chest 2005; 128: 2068-75
9. Curkendall S, de Luise C, Jones J, et al. Cardiovascular disease in patients with chronic obstructive pulmonary disease. Saskatchewan, Canada. Ann Epidemiol 2006; 16: 63-70
10. National Hospital Discharge Survey [online]. Available from URL: http://www.cdc.gov/nchs/data/ad/ad319.pdf [Accessed 2005 Jul 7]
11. Quam L, Ellis LBM, Venus P, et al. Using claims data for epidemiologic research: the concordance of claims-based criteria with the medical record and patient survey for identifying a hypertensive population. Med Care 1993; 131: 498-507
12. Lanza L, Dreyer NA, Schultz NJ, et al. Use of insurance claims in epidemiologic research: identification of peptic ulcers, GI bleeding, pancreatitis, hepatitis and renal disease. Pharmacoepidemiol Drug Saf 1995; 4: 239-48
13. Sands BE, Duh MS, Cali C, et al. Algorithms to identify colonic ischemia, complications of constipation and irritable bowel syndrome in medical claims data: development and validation. Pharmacoepidemiol Drug Saf 2006; 15: 47-56
14. Psaty BM, Boineau R, Kuller LH, et al. The potential costs of upcoding for heart failure in the United States. Am J Cardiol 1999; 84: 108-9
15. Silverman E, Skinner J. Medicare upcoding and hospital ownership. J Health Econ 2004; 23 (2): 369-89
16. West SL, Strom BL, Poole C. Validity of pharmacoepidemiologic drug and diagnosis data. In: Strom BL, editor. Pharmacoepidemiology. 4th ed. New York (NY): John Wiley & Sons Ltd, 2005: 709-66
17. Rancourt J. UnitedHealth's i3 unit tapped by FDA to provide postmarketing surveillance. FDC Reports, Inc. The Pink Sheet Daily 2005 Sep 28. No. 6
18. Rothman KJ, Greenland S, editors. Modern epidemiology. 2nd ed. Philadelphia (PA): Lippincott-Raven, 1998: 130-4

Correspondence and offprints: Dr *Stephan F. Lanes*, Boehringer Ingelheim Pharmaceuticals Inc., 900 Ridgebury Road, PO Box 368, Ridgefield, CT 06877-0368, USA.

E-mail: slanes@rdg.boehringer-ingelheim.com