

Commentary: Establishing Standards for the Utility of Administrative Claims Data

Sheldon M. Retchin and David J. Ballard

The utility of administrative healthcare claims data for evaluating the delivery of health services has been well established. Over the past 25 years, the analysis of claims data has revealed unexplained practice variations in neighboring geographic markets (Wennberg et al. 1989), uncovered differential access to appropriate care by race (Desch, Penberthy, Newschaffer, et al. 1996) and gender (Ayanian and Epstein 1991), and detected opportunities to improve quality of care (Lohr 1990; Weiner et al. 1990). Furthermore, administrative claims have also been used to estimate the incidence of disease (McBean, Warren, and Babish 1994) and the outcomes of surgical procedures (Lubitz et al. 1993; Mitchell, Bubolz, Paul, et al. 1994). Finally, these data have been useful in evaluating the dissemination of new innovations into practice.

ADVANTAGES AND DISADVANTAGES OF ADMINISTRATIVE CLAIMS

Administrative claims are collected as a part of routine clinical service delivery, conventionally as a requirement for reimbursement. Because claims are so uniformly generated, they can offer a relatively unique, population-based data source when appropriate denominator information can be obtained (White 1997). Another advantage is that the data are relatively inexpensive to gather and frequently offer fewer restrictions in terms of sample size than do data from alternative sources. Further, administrative data have been

Address correspondence and requests for reprints to Sheldon M. Retchin, M.D., M.S.P.H., Box 980270, Richmond, VA 23298-0270. Dr. Retchin is Professor of Internal Medicine and Associate Vice President, Virginia Commonwealth University. David J. Ballard, M.D., Ph.D. is Professor of Medicine and Director, Center for Clinical Evaluation Sciences; Emory University, and President of the Kerr L. White Institute for Health Services Research in Decatur, GA.

collected for more than two decades, and thus can often provide insights into trends in healthcare and clinical outcomes (Platt, Svenson, and Woodhead 1993). Despite the advantages of administrative data sources, there are also problems inherent in the use of information that is collected for purposes other than research (Mitchell, Bubolz, Paul, et al. 1994). The information obtained from the data may be inaccurate, or even spurious. For instance, diagnostic information is often required for remuneration, and more complex diagnoses can be rewarded with higher payments. This can result in "upcoding" or "gaming." Hsia et al. reported that more than 20 percent of the diagnostic assignments at admission through the prospective payment system needed modification after the medical records were reviewed (Hsia, Krushat, Fagan, et al. 1988). These inaccuracies can bias research efforts that compare practice patterns of hospitals or other providers. Moreover, when documentation of care is deemed nonessential for administrative/reimbursement purposes, some important elements may be incomplete. And since the generation of claims data usually follows the utilization patterns of clinical care, disproportionate details are more readily accessible for inpatient episodes than for outpatient care. Thus, since administrative claims are likely to continue to be an important source of information on health delivery, epidemiologic surveillance, and clinical outcomes, it will be vital to establish criteria, or parameters of performance, that determine their precision.

ESTABLISHING STANDARDS FOR THE USE OF ADMINISTRATIVE CLAIMS

Given that the promise of administrative data sources is so alluring and yet the problems are so troubling, studies that explore the accuracy of claims data are especially meaningful. For these reasons, undertakings such as those by Horner, Sloane, and Kahn, in this issue of *Health Services Research*, are integral to future applications. In their study, the authors used hospital medical records to determine the accuracy of administrative claims for classifying the clinical severity of patients admitted to hospitals for new strokes (Horner, Sloane, and Kahn 1998). They used these data to demonstrate that mechanical ventilation on the first day of hospitalization is a useful proxy for coma, as well as a prognostic indicator for stroke. The benefit of this study could be a surrogate measure of stroke severity that may be used to evaluate the relevance of practice variations or to examine outcomes of care.

Nonetheless, this article also demonstrates the imperfections of administrative data, and the reasons why standards for their validation are so

important. Since administrative databases are generated for purposes other than to support clinical care and health services research, the findings of Horner, Sloane, and Kahn remind us that these data require verification when used to describe and evaluate clinical care. Several issues should be considered in substantiating the accuracy of administrative data.

First, studies that attempt to verify the use of administrative data depend on the accuracy and timeliness of the claims. For instance, since the data in the study by Horner et al. were collected on the care delivered during two time periods, now ten years old, there might be concerns over changes in availability of information since that time. Capitation arrangements are increasingly made between health plans and providers (Robinson and Casalino 1995), and these agreements offer little incentive for the submission of accurate claims data. Although there is some promise that health plans will collect data to comply with the requirements of accrediting organizations, many have expressed anxiety over the precision of "dummy claims" submitted by hospitals with managed care contracts. While the majority of provider remuneration remains fee-for-service (Gold, Hurley, Lake, et al. 1995), the ability of health plans to produce useful data remains limited. In fact, this limitation has resulted in recent proposals requiring health maintenance organizations to submit encounter-level data in exchange for capitated premiums for Medicare (Welch and Welch 1995). Thus, a careful description of health plan influences in sampling administrative claims should be provided.

Second, depending on the purpose of administrative claims, clinical data from medical records or abstracts should be employed to validate measures to be used for research purposes. Alternatively, prospective data collection efforts, either directly from patients or from other sources (e.g., providers), may be used. The former approach is the method used by Horner et al., whereby clinical data collected from medical records were used to validate the use of mechanical ventilation for identifying the presence of coma among patients with acute strokes. Nonetheless, even though the same administrative database may be used for different measures, and across different patient populations, it should not be assumed that a prior validation study is applicable. The generalizability to other settings and diagnoses may be limited. Thus, mechanical ventilation as a proxy indicator of coma may be acceptable for stroke patients, but it would not be as successful as a proxy measure for coma among patients with Guillain-Barré syndrome. Therefore, further verification would be required to test this approach as it is applied to other conditions, and perhaps in other populations.

Third, it should be recognized that the utility of administrative data is not static. The purpose and quality of coding will likely change over time.

The use of constructs derived from administrative claims as proxy measures for representing clinical phenomena will often be coupled to standards of care (e.g., the use of mechanical ventilation as a quality of care indicator for stroke patients with coma). This is probably unavoidable. However, variations in practice are well known, and standards of care change. There is also the potential that hospitals will become more proficient with coding. For example, the agreement between abstracted medical record charts and the presence of cancer from administrative claims increased from 57 percent in 1977 to 84 percent in 1985 (Fisher, Whaley, Krushat, et al. 1992). In contrast, the positive predictive value of hospital discharge abstracts with the discharge diagnosis ICD codes 430-438.9 for identifying stroke events did not improve from 1970 through 1989 (Leibson et al. 1994). For these reasons, validation efforts with administrative data should be updated periodically.

Fourth, analysis of specific elements contained in administrative data should be avoided if gaming, or upcoding is probable. This will likely be peculiar to specific measures or variables, and may necessarily emphasize the use of procedures or interventions. Thus, the accuracy of the use of a code for mechanical ventilation for the purpose of classifying severity in stroke, or the diagnosis of cancer, would be less disputable than for a diagnosis such as congestive heart failure. For the former, chart audit would verify the diagnosis for a specific procedure (i.e., mechanical ventilation) or pathologic reading (e.g., cancer); however, verification of congestive heart failure via chart audit would be more problematic (Green and Wintfeld 1993).

Finally, conventional methods of agreement or validation should be used when evaluating the utility of administrative claims for research purposes. Where no gold standard exists, kappa values may be used to assess agreement between two or more data sources. When there is consensus regarding a gold standard (e.g., medical record), and validation is the aim, traditional measures of accuracy and precision should be employed. Although the kappa coefficients reported by Horner et al. suggested that most of the cases in their sample were stroke patients, others have questioned the precision of this diagnosis, even on hospitalized patients. In the Rochester Epidemiology Project (Melton 1996), only 76 percent of residents in Rochester, Minnesota who were hospitalized with an incident stroke were actually assigned a principal diagnosis code of cerebrovascular disease in the hospital abstract (Leibson et al. 1994). Moreover, for the Horner et al. study, if the medical record was considered a gold standard, it would have had exceptionally high positive (93.2 percent) and negative (80.4 percent) predictive values. And yet, in the Rochester study, the positive predictive value of a principal diagnosis of stroke

(ICD-9 430–438.9) was only 60 percent. This serves to emphasize that the precision of the diagnosis of stroke, even among hospitalized patients, is not irrefutable (Leibson et al. 1994). Also, predictive values are sensitive to the prevalence of the “condition,” in this case the use of mechanical ventilation. Therefore, the sensitivity of using this proxy measure for the presence of coma would be limited (17.3 percent in this study), reflecting the low occurrence (4 percent) of mechanical ventilation in the cohort evaluated by Horner et al. These discrepancies also emphasize the need to use traditional validation measures when evaluating the accuracy of proxy measures from administrative data sources, so that the findings can be compared in different populations.

FUTURE UTILITY OF ADMINISTRATIVE DATA

The convenience, accessibility, and unobtrusiveness of administrative data for research purposes are features that cannot be ignored. The rapid increase in the use of claims for research in recent years is evidence that their value has been widely recognized. However, future health services research applications of administrative data could be enhanced or jeopardized by a lack of agreement on standards for their utility. Further endeavors, such as those by Horner, Sloane, and Kahn, are needed to validate the utility of administrative data as proxy clinical measures before their full benefit can be realized.

REFERENCES

- Ayanian, J. Z., and A. M. Epstein. 1991. “Differences in the Use of Procedures Between Women and Men Hospitalized for Coronary Artery Disease.” *The New England Journal of Medicine* 325 (4): 221–25.
- Desch, C. E., L. Penberthy, C. J. Newschaffer, B. E. Hillner, M. Whittemore, D. McClish, T. J. Smith, and S. M. Retchin. 1996. “Factors That Determine the Treatment for Local and Regional Prostate Cancer.” *Medical Care* 34 (2): 152–62.
- Fisher, E. S., F. S. Whaley, W. M. Krushat, D. J. Malenka, C. Fleming, J. A. Baron, and D. C. Hsia. 1992. “The Accuracy of Medicare’s Hospital Claims Data: Progress Has Been Made, but Problems Remain.” *American Journal of Public Health* 82 (February): 243–48.
- Gold, M. R., R. Hurley, T. Lake, T. Ensor, and R. Berenson. 1995. “A National Survey of the Arrangements Managed Care Plans Make with Physicians.” *The New England Journal of Medicine* 333 (25): 1678–83.

- Green, J., and N. Wintfeld. 1993. "How Accurate Are Hospital Discharge Data for Evaluating Effectiveness of Care?" *Medical Care* 31 (8): 719-31.
- Horner, R. D., R. J. Sloane, and K. L. Kahn. 1998. "Is Use of Mechanical Ventilation a Reasonable Proxy Indicator for Coma Among Medicare Patients Hospitalized for Acute Stroke?" *Health Services Research* 32 (6): 843-62.
- Hsia, D. C., W. M. Krushat, A. B. Fagan, J. A. Tebbutt, and R. P. Kusserow. 1988. "Accuracy of Diagnostic Coding for Medicare Patients under the Prospective Payment System." *The New England Journal of Medicine* 318 (6): 352-55.
- Leibson, C. L., J. M. Naessens, R. D. Brown, and J. P. Whisnant. 1994. "Accuracy of Hospital Discharge Abstracts for Identifying Stroke." *Stroke* 25 (12): 2348-55.
- Lohr, K. N. 1990. "Use of Insurance Claims Data in Measuring Quality of Care." *International Journal of Technology Assessment in Health Care* 6 (2): 263-71.
- Lubitz, J. D., M. E. Gornick, R. M. Mentnech, and F. D. Loop. 1993. "Rehospitalizations after Coronary Revascularization among Medicare Beneficiaries." *American Journal of Cardiology* 72 (1): 26-30.
- McBean, A. M., J. L. Warren, and J. D. Babish. 1994. "Measuring the Incidence of Cancer in Elderly Americans Using Medicare Claims Data." *Cancer* 73 (9): 2417-25.
- Melton, L. J., III. 1996. "History of the Rochester Epidemiology Project." *Mayo Clinic Proceedings* 71 (3): 266-74.
- Mitchell, J. B., T. Bubolz, J. E. Paul, C. L. Pashos, J. J. Escarce, L. H. Muhlbaier, J. M. Wiesman, W. W. Young, R. S. Epstein, and J. C. Javitt. 1994. "Using Medicare Claims for Outcomes Research." *Medical Care* 32 (7) (Supplement): JS38-JS51.
- Platt, G. H., L. W. Svenson, and S. E. Woodhead. 1993. "Coronary Artery Bypass Grafting in Alberta from 1984 to 1989." *Canadian Journal of Cardiology* 9 (7): 621-24.
- Robinson, J. C., and L. P. Casalino. 1995. "The Growth of Medical Groups Paid Through Capitation in California." *The New England Journal of Medicine* 333 (25): 1684-87.
- Weiner, J. P., N. R. Powe, D. M. Steinwachs, and G. Dent. 1990. "Applying Insurance Claims Data to Assess Quality of Care: A Compilation of Potential Indicators." *Quality Review Bulletin* 16 (12): 424-38.
- Welch, W. P., and H. G. Welch. 1995. "Fee-for-Data: A Strategy to Open the HMO Black Box." *Health Affairs* 14 (4): 104-16.
- Wennberg, J. E., J. L. Freeman, R. M. Shelton, and T. A. Bubolz. 1989. "Hospital Use and Mortality among Medicare Beneficiaries in Boston and New Haven." *The New England Journal of Medicine* 321 (17): 1168-73.
- White, K. L. 1997. "The Ecology of Medical Care: Origins and Implications for Population-Based Healthcare Research." *Health Services Research* 32 (1): 11-21.