Guide to Statistics and Methods

Practical Guide to Surgical Data Sets: Medicare Claims Data

Amir A. Ghaferi, MD, MS; Justin B. Dimick, MD, MPH

Introduction

The Centers for Medicare and Medicaid Services (CMS) administers Medicare, the primary US health insurance program for people aged 65 years and older and people who qualify for Social Security Administration disability benefits (**Box**). Medicare includes Part A, which is hospital insurance; Part B, medical insurance; Part C, Medicare Advantage (private health insurance approved by the CMS and paid on a per-capita basis); and Part D, prescription drug coverage. The CMS maintains and makes several data files available for purchase (https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/FilesForOrderGenInfo/index.html). Because the Medicare Advantage data are administered by private insurance, these claims are unavailable. However, national fee-for-service Medicare files are available and represent approximately 70% of beneficiaries.

Medicare data represent claims submitted to the CMS for reimbursement of services rendered. The Medicare data set has very little missing data because accurate claims are necessary for hospital and physician payments.

Pros and Cons of Medicare Data

Several features make this data set a useful research tool. First, specific demographic data are included (eg, age, birthdate, sex, race/ethnicity, and place of residence). Second, these data can be linked to other CMS data sets on health care utilization, insurance enrollment, and clinician characteristics. Third, the data cover nearly 70% of adults aged 65 years and older, making fee-for-service Medicare data a rich source of utilization and outcomes data and allowing for subgroup analyses without decreased statistical power. Fourth, the data can be linked to non-CMS data, such as the US Census, cancer registries (eg, the Surveillance, Epidemiology, and End Results Program; Medicare), other government insurance programs (eg, Medicaid), the Social Security death index, and clinician information (ie, American Hospital Association data). Fifth, patients can be tracked across episodes of care, which permits longitudinal evaluations of outcomes and health care utilization. Finally, Medicare data files are a cost-effective way to assess a large patient population across multiple health care settings.

However, there are important limitations when using Medicare data for research. First, it only includes a diagnosis documented via the *International Classification of Diseases, Ninth Revision (ICD-9)* or *ICD-10* codes (ie, hypertension, depression, diabetes). This can prove difficult when assessing surgical complications. Iezzoni et al[1] identified the most appropriate *ICD* codes to measure comorbidities and complications, but these are imperfect. Second, there is no physiological or biochemical patient information, such as vital signs, laboratory test results, and pathology results. Third, there are no timestamps during a hospital stay. This can limit the study of how care progresses or when events and complications occur during a hospitalization. Fourth, evaluation of outpatient utilization patterns can be limited by the lack of data on uncovered services or benefits and managed care enrollee information. Finally, derivation from billing data limits reliability. This is especially important in patients undergoing surgery, for whom comorbidity and severity of illness may be inconsistently documented.

Box. Attributes of Medicare Claim Data

1. Medicare data are an excellent national representation of a large proportion of the older adult population.
2. While a cost-effective way of evaluating a large population, securing independent funding to purchase the data is highly recommended.
3. The data sets available from the Centers for Medicare and Medicaid Services are suitable for linkages to several existing data sets (ie, American Heart Association, US Census, and others).
4. Data can be tracked longitudinally across episodes of care, making this a uniquely positioned dataset to study long-term outcomes in surgical patients.
5. Several advanced statistical methods can increase the robustness of inferences made using this data; inclusion of experienced methodologists in research is highly recommended.

Potential Avenues of Research

Medicare data can provide valuable insights to several topical areas of surgical research. Three commonly studied categories are health policy evaluation, comparative effectiveness research, and outcome variations. The **Table** summarizes these 3 themes and methodologies.

Table. Common Research Uses for Medicare Data and Statistical Approaches for Addressing Common Methodological Problems

Major Research Theme	Methodological Problem	Statistical Methods
Comparative effectiveness research	Adjusting for selection bias (eg, differences in characteristics between treated and control groups)	Multivariate regression, instrumental variable analysis, and propensity score analysis
Health policy evaluation	Adjusting for background time trends (eg, improvements in outcomes over time)	Difference-in-differences analysis
Understanding variation across clinicians	Risk adjustment (eg, accounting for differences in patient characteristics across clinicians); accounting for clustering within clinicians and adjusting for reliability	Multivariate regression and hierarchical modeling

Comparative Effectiveness Research

Comparative effectiveness research is the direct comparison of health care interventions to determine which works best, for whom, and in what circumstances. Although randomized clinical trials are excellent at identifying the efficacy of a procedure or treatment, these studies are limited by strict inclusion criteria and short-term follow-up. Large claims databases facilitate assessments in real-world settings with broader groups of patients and clinicians. For example, after 2 large national trials demonstrated the efficacy of carotid endarterectomy, a classic comparative effectiveness study[2] evaluated the real-world effectiveness of the procedure outside the confines of a clinical trial. Using multivariate logistic regression, Wennberg et al[2] found perioperative mortality after carotid endarterectomy in patients receiving Medicare to be markedly higher than in a control group. This raised concerns about translating randomized clinical trials' efficacy into effectiveness in everyday practice. Researchers interested in using Medicare data to study comparative effectiveness should develop expertise or seek collaborators experienced in using advanced methods for causal inference, such as instrumental variable analysis[3] and propensity score matching.[4]

Health Policy Evaluation

Health policy shapes many aspects of our environment, including payment, performance measurement, and training. However, to our knowledge, there is little evidence that policies have the desired effect of improving outcomes or reducing costs. Medicare data provide the use of rigorous methods on a large cohort of patients and clinicians to examine broad policy implications, intended and some unintended. Dimick et al[5] evaluated the implications of a CMS national coverage decision restricting bariatric surgery to centers of excellence. However, studies supporting this well-intentioned policy lacked control groups. Using longitudinal Medicare claims data and sophisticated statistical methods, this study identified no difference in outcomes before and after policy implementation, leading to a reconsideration and ultimately a reversal of that coverage decision. The specific method used was difference-in-differences, an econometric technique that accounts for background trends in outcomes and is common in health care policy evaluations.

Understanding Variation

Developing a better understanding of the magnitude of variation and its associations with measurable characteristics is a valuable way to gain insight into surgical practice and levers for change. These evaluations lend themselves to new policies to reduce variation in surgical care. Ghaferi et al[6] examined drivers of variation in mortality rates with high-risk surgery in patients receiving Medicare benefits. Mortality rates varied significantly, with a nearly 2.5-fold difference between the best and worst hospitals. Most interestingly, complication rates were similar across hospitals, but the rate at which patients were rescued from complications were much higher at hospitals with low mortality. These findings helped guide surgical quality improvement efforts toward reducing failure to rescue rates by focusing on timely and effective management of postoperative complications. The statistical methods used in this study included multivariate logistic regression to adjust for confounding patient variables and reliability adjustment to account for random chance as a driver of outcomes.[7]

Where to Find More Information

Medicare data are useful for studying comparative effectiveness of procedures, health care policy, and outcome variations. However, it is important to frame questions carefully and use appropriate methods to ensure scientific rigor, as each of the cited studies have done.

REFERENCES

1. Iezzoni LI, Daley J, Heeren T, et al. Identifying complications of care using administrative data. *Med Care*. 1994;32(7):700-715.

2. Wennberg DE, Lucas FL, Birkmeyer JD, Bredenberg CE, Fisher ES. Variation in carotid endarterectomy mortality in the Medicare population: trial hospitals, volume, and patient characteristics. *JAMA*. 1998;279(16):1278-1281.

3. Tan HJ, Norton EC, Ye Z, Hafez KS, Gore JL, Miller DC. Long-term survival following partial vs radical nephrectomy among older patients with early-stage kidney cancer. *JAMA*. 2012;307(15):1629-1635.

4. Haukoos JS, Lewis RJ. The Propensity Score. *JAMA*. 2015;314(15):1637-1638.

5. Dimick JB, Nicholas LH, Ryan AM, Thumma JR, Birkmeyer JD. Bariatric surgery complications before vs after implementation of a national policy restricting coverage to centers of excellence. *JAMA*. 2013;309(8):792-799.

6. Ghaferi AA, Birkmeyer JD, Dimick JB. Complications, failure to rescue, and mortality with major inpatient surgery in Medicare patients. *Ann Surg*. 2009;250(6):1029-1034.

7. Dimick JB, Ghaferi AA, Osborne NH, Ko CY, Hall BL. Reliability adjustment for reporting hospital outcomes with surgery. *Ann Surg*. 2012;255(4):703-707.