Invited commentary: databases for surgical health services research: Clinformatics Data Mart ☆

Introduction

Administrative claims have been a vital source of data for researchers studying healthcare utilization, spending, and outcomes in the United States. Claims data comprise member enrollment information and billing data submitted by hospitals, physicians, pharmacies, and other providers to payers for reimbursement. Almost every encounter a patient has with the healthcare system generates a claim. Furthermore, these data are relatively standardized because diagnoses, procedures, and medications are recorded, using specific codes. Research using claims data has become more prevalent because of the large sample size and standardized format. This review will focus on Clinformatics Data Mart (OptumInsight, Eden Prairie, MN). We will specifically discuss the types of data available in this dataset, its strengths and limitations, and how these data can serve the needs of healthcare researchers to provide more insight and promote meaningful change.

What is Clinformatics Data Mart?

The Clinformatics Data Mart is from Optum and contains de-identified patient-level claims data from a national insurance provider for enrollees with both medical and pharmacy benefits. The database includes more than 10 years of commercial claims (2004–2016) and claims from Medicare Advantage plans (2006–2016). The Medicare claims account for approximately 25% of Medicare Advantage enrollees in the United States. The Clinformatics Data Mart database provides 2 distinct datasets:

- Core datasets: These datasets comprise member enrollment, inpatient, physician and facility (medical) claims, pharmacy claims, and laboratory test results data. An analytic dataset can be constructed by integrating any or all of the core datasets.
- Extended datasets: These datasets include socioeconomic status, date of death, geographic information, and vision and youth

data. To protect patient confidentiality, only one of the extended datasets can be linked to the core datasets.

Strengths and Limitations

Strengths

The availability of data for a large number of enrollees throughout multiple years in the Clinformatics Data Mart is attractive to researchers for several reasons. First, these data make it possible to investigate trends over time. This offers a longitudinal view of a patient's overall health. For example, for patients undergoing bariatric surgery, researchers can track obesity-related comorbidities up to 5 years after surgery.[1] The large sample size provides substantial statistical power for a study, which is particularly helpful when complex statistical modeling is required. In addition, the dataset comprises enrollees and their families from various regions across the country. Because of this, patient profiles are heterogeneous, which helps to avoid geographic biases and leads to more generalizable conclusions. Last, the large size suggests that the data are likely to capture rare events or diseases, such as uncommon procedures, outcomes, or complications.[2]

Another important strength of this dataset is that it contains data on all healthcare services that are provided to patients covered by UnitedHealthcare, regardless of where or from whom they received their care. As a result, it is feasible to track utilization and outcomes of care, as well as the healthcare spending for these services. Furthermore, data on the family members of the primary enrollee are also available. This allows researchers to explore the role of family-level variables on the outcome of interest.

The Clinformatics Data Mart also provides encrypted data on the healthcare providers who render services to patients. This information allows for comparison of practice patterns between physicians and among various specialty groups. The effect of socioeconomic status or geographic variations on healthcare utilization can also be studied.

Limitations

Despite its many strengths, the Clinformatics Data Mart has several limitations. First, undercoding and miscoding do occur, as common limitations across many claims sources regardless of data provider, because diagnoses, procedures, and medications may be coded to maximize reimbursement.[3,4] As a result, trends must be

interpreted with caution as they may be a result of coding practices rather than clinical events. Another key limitation is that some clinical data such as blood loss, length of surgery, laboratory results, medications administered during a hospital admission, and complication severity are not available in the dataset. In addition, clinical notes are not available in claims data, making it difficult to obtain the complete medical history of a patient. No patient-reported data are included, such as improvement in pain or function after surgery, which are critical when considering surgical outcomes. Another key limitation is that the dataset is limited to a single commercial insurer. Therefore, data are not available for any services provided out-of-network or paid for by other payers. In addition, findings may not be generalizable to other populations, such as Medicare fee-for-service patients. Finally, because patient data are de-identified, it is not possible to link this dataset to other databases, such as clinical registries.

What Questions Can the Dataset Answer?

The Clinformatics Data Mart enables researchers to investigate healthcare utilization, spending, and outcomes (Table). For example, Maradit Kremers et al[5] studied the utilization of hip arthroscopy and total hip arthroplasty over time. The data can also be used to study rare events, such as postoperative venous thromboembolism.[2] Prolonged opioid use among surgical patients captured in pharmacy claim data can be linked with patient and clinical characteristics.[6] Pharmacy claims provide detailed information for identifying the optimal quantity of initial postoperative opioid prescriptions without additional refill.[7] The DataMart also provides a standardized price for medical services, which adjusts for changes in payments based on local factors, such as labor costs. This allows more accurate study and comparison of spending for healthcare services. Furthermore, researchers can evaluate the impact of socioeconomic factors, such as ethnicity, income, education and marital status.

Summary

In conclusion, the Clinformatics Data Mart is a large claims-based dataset that uses national data from a single commercial payer. Its major strengths include its size, the granularity of data, and its standardized format. It is limited by the potential for miscoding, a lack of key clinical data, and its inability to generalize findings to Medicare patients or uninsured patients. Despite these limitations, it is a valuable dataset that can be effectively used to study healthcare utilization, spending, and outcomes in surgical patients.

Vidhya Gunaseelan, Brooke Kenney, Jay Soong-Jin Lee, Hsou Mei Hu
Department of Surgery, University of Michigan Medical School, Ann Arbor, USA
Michigan Opioid Prescribing Engagement Network (Michigan OPEN), Ann Arbor, USA

* Corresponding author. Department of Surgery, University of Michigan, School of Medicine, 1500 E Medical Center Drive, Ann Arbor, MI 48109, USA. (H.M. Hu).
E-mail address: hsoumeih@med.umich.edu (H.M. Hu)

References

1 Patkar A, Fegelman E, R Kashyap S, et al. Assessing the real-world effect of laparoscopic bariatric surgery on the management of obesity-related comorbidities: a retrospective matched cohort study using a US Claims Database. *Diabetes Obes Metab.* 2017;19:181–188.
2 Brady MT, Patts GJ, Rosen A, et al. Postoperative venous thromboembolism in patients undergoing abdominal surgery for IBD: a common but rarely addressed problem. *Dis Colon Rectum.* 2017;60:61–67.
3 Iezzoni LI, Daley J, Heeren T, et al. Identifying complications of care using administrative data. *Med Care.* 1994;32:700–715.
4 Davenport DL, Holsapple CW, Conigliaro J. Assessing surgical quality using administrative and clinical data sets: a direct comparison of the University HealthSystem Consortium Clinical Database and the National Surgical Quality Improvement Program data set. *Am J Med Qual.* 2009;24:395–402.
5 Maradit Kremers H, Schilz SR, Van Houten HK, et al. Trends in utilization and outcomes of hip arthroscopy in the United States between 2005 and 2013. *J Arthroplasty.* 2017;32:750–755.
6 Brummett CM, Waljee JF, Goesling J, et al. New persistent opioid use after minor and major surgical procedures in US adults. *JAMA Surg.* 2017;152:e170504.
7 Sekhri S, Arora NS, Cottrell H, et al. Probability of opioid prescription refilling after surgery: Does initial prescription dose matter? *Ann Surg.* 2017 Epub ahead of print. doi:10.1097/SLA.0000000000002308.

Table
Examples of work based on clinformatics data mart database

Authors	Study population	Research question	Outcome(s)
Patkar et al[1]	Patients undergoing laparoscopic bariatric surgery between 2006 and 2013 who had a diagnosis code of obesity and had at least one of obesity-related comorbidity (type 2 diabetes, dyslipidaemia, hypertension, depression, or sleep apnea).	Did laparoscopic bariatric surgery improve the management of obesity-related comorbidities?	– Number of medication classes used to treat type 2 diabetes, hypertension, and dyslipidaemia. – Percentage of patients not needing medications for these comorbidities.
Brady et.al[2]	Patients aged 18 years and older undergoing surgery for Crohn's disease or ulcerative Colitis between 2004 and 2013.	What percent of patients had postdischarge venous thromboembolism chemical prophylaxis? What was postdischarge 90-day venous thromboembolism rate? What were the predictors of postdischarge venous thromboembolism?	– Postdischarge venous thromboembolism Prophylaxis. – 90-day postdischarge thromboembolic events.
Maradit Kremers et al[5]	Aged 18–64 years undergoing a hip arthroscopy procedure between 2005 and 2013.	How many of patients had a subsequent hip arthroscopy and/or total hip arthroplasty? What was the trend over time?	– Undergoing subsequent hip arthroscopy. – Total hip arthroplasty rate.
Brummett et al[6]	Aged 18–64 years, opioid-naïve patients undergoing major or minor surgery who filled an opioid prescription for the surgery in 2013–2014.	What were the rates of new persistent use? What were risk factors (patient and clinical characteristics) for new persistent use?	– New persistent opioid use, defined as an opioid prescription fulfillment between 90 and 180 days after the surgical procedure. – Refill opioid prescription within 30 days after surgery.
Sekhri et al[7]	Aged 18–64 years, opioid-naïve patients undergoing major or minor surgery who filled an opioid prescription for the surgery in 2013–2014.	Was there correlation between probability of postoperative opioid prescription refills and the amount of opioid prescribed? Did a greater initial prescription yield a lower probability of refill?	