Working with Existing Databases

Melissa Murphy, MD, MPH¹ Karim Alavi, MD, MPH¹ Justin Maykel, MD¹

¹ Department of Colon and Rectal Surgery, University of Massachusetts Medical School, Worcester, Massachusetts

Address for correspondence Justin Maykel, MD, Department of Surgery, University of Massachusetts Medical School, 67 Belmont St., Worcester, MA 01605 (e-mail: Justin.Maykel@umassmemorial.org).

Clin Colon Rectal Surg 2013;26:5–11.

Abstract

Keywords

- outcomes
- databases
- administrative
- research

Outcomes research has established itself as an integral part of surgical research as physicians and hospitals are increasingly required to demonstrate attainment of performance markers and surgical safety indicators. Large-volume and clinical and administrative databases are used to study regional practice pattern variations, health care disparities, and resource utilization. Understanding the unique strengths and limitations of these large databases is critical to performing quality surgical outcomes research. In the current work, we review the currently available large-volume databases including selection processes, modes of analyses, data application, and limitations.

Objectives: On completion of the article, the reader should be able to describe the differences between large volume administrative and clinical databases and to discuss the limitations of using administrative databases for outcomes research.

Surgical outcomes research has grown exponentially in the past decade with increased utilization of large-volume administrative and clinical databases. Surgeon, hospital, and regional outcomes are increasingly evaluated with comparison of both patient outcomes (morbidity and mortality) and resource utilization (cost, intensive care unit admissions). External pressures have further fueled outcomes research with the American Board of Surgery requiring surgeons to monitor their own performance,¹ the Joint Commission urging hospitals to demonstrate key surgical safety indicators,² and payers implementing pay-for-performance programs.³

Within the field of colon and rectal surgery, large-volume databases are increasingly utilized to study outcomes.^{4,5} For both outcomes analysis as well as research, these provide multiple potential advantages. Derivation of data from large, heterogeneous populations may result in decreased selection bias and increased study generalizability.^{6,7} New surgical interventions⁸ and assessment of quality improvement measures may be examined.^{9,10} In the era of escalating health care costs, database research lends itself to the assessment of the cost and effectiveness of surgical treatments¹¹ while permitting investigation of the role of specialty provider and institution characteristics on resource utilization and patient

outcomes. Variations at the regional, state, and national level can be assessed and risk models can be created and validated.¹²

On the other hand, as large-volume databases become increasingly available for research purposes, a clear understanding of the appropriate applications and limitations is necessary. The nuances and complexities of individual databases can be confusing. In this section, we aim to discuss the large-volume databases currently available for outcomes research, outlining database selection, modes of analyses, and inherent limitations.

Study Design

Clinical research includes the traditional prospective clinical trials, cohort studies, and case-control studies, while outcomes research has now solidified itself as a separate class of clinical research. Subtypes of surgical outcomes research include geographic variations of surgical procedures, volume outcomes analyses, treatment disparities (racial/economic/age-related), trend analyses, cost-effectiveness studies, and surgical quality/risk adjustment.

Appropriate database selection requires a thorough understanding of the individual database, including unique strengths and limitations. The utility of administrative data for assessing surgical quality depends on the measures being studied and how well they are captured in administrative and clinical records. In addition, the initial study design and hypothesis generation is a critical phase to the validity and

success of a study. Chang et al proposed a “research protocol” to facilitate research using large-volume databases.¹³ Key elements in the design phase include a clear definition of the study population (i.e., rectal cancer patients, over age 65), subset definition (i.e., minorities, elderly, stage), outcomes variable(s) (i.e., mortality, complications, length of stay), definition of the primary comparison (i.e., Are elderly rectal cancer patients less likely to undergo sphincter preserving surgery?), and identification of covariates/confounders (i.e., age, gender, race, socioeconomic status). Selection of the appropriate database follows the study design phase.

Databases

Generally, large-volume databases can be broadly categorized as either administrative or clinical based on where the data are derived (►Fig. 1). Administrative databases are largely based on billing information while clinical databases are populated with defined patient information. The geographic catchment areas of individual databases may also vary. They may be national, state-based, or limited to specific geographic areas. All patients in a particular catchment area may be included, or limited to a representative sample. The population within the database may be limited additionally by specific restrictions including age (Medicare), military status (veterans), or diagnosis (cancer registries).

Administrative

Administrative databases were not originally designed for clinical research, but rather to track billing for hospitals, providers, and procedures. Administrative data are derived from two sources: requests to insurers for health care payments and claims for clinical services and therapies. Compared with performing a national prospective study, access to administrative data are readily available and relatively inexpensive.

Fig. 1 Large-volume administrative and clinical databases. ACTUR, automated Central Tumor Registry; CanCORS, Cancer Care Outcomes and Research Consortium; HCUP, Healthcare Cost and Utilization Project; KID, Kids Inpatient Database; NCDB, National Cancer Database; NIS, Nationwide Inpatient Sample; NSQIP, National Surgical Quality Improvement Program; NTDB, National Trauma Data Bank; SEER, The Surveillance, Epidemiology, and End Results Program; SID, State Inpatient Database; UHC, University Healthsystem Consortium.

Utilizing international classification of disease clinical medicine (ICD-9-CM) codes for diagnoses and current procedural terminology (CPT) codes, administrative databases provide a mechanism to look at population-based data. The large number of patients allows for sufficient powering of research studies. Topics such as operative trends may be assessed using longitudinal databases spanning several years. Additional outcomes including length of stay and hospital cost may be assessed with large administrative databases.

Although providing a wealth of information, administrative databases are not without limitations. Coders are relied upon for the accuracy of the data and validity of medical diagnoses.^{14,15} Patient-level clinical information is not available (stage of disease, tumor characteristics, preoperative performance status). Furthermore, specifics of the operation (i.e., conversion from laparoscopic to open, intraoperative findings, etc.) are also not tracked. Morbidity information is limited to ICD-9 codes and some ambiguity exists whether it is a true morbidity or complication.¹⁶ Mortality may be assessed within the administrative databases or through linkage with clinical databases, yet there is limited available data on nonfatal and subjective patient outcomes (quality of life, patient satisfaction).

Caution should be taken regarding the generalizability of study results derived from administrative databases. Although 95% of the eligible U.S. elderly population are enrolled in Medicare coverage (Part A 98%, Part B 96%), variations by age, sex, and race are not recorded.¹⁷ Medicare Part C (managed care) covers nearly 20% of Medicare beneficiaries, but the plans are not required to submit claims for research purposes. Additionally, studies performed in insured populations may not be applicable to uninsured populations as their disease patterns and health care needs may differ.¹⁷

Medicare and Medicaid

The United States (US) Centers for Medicare and Medicaid Services (CMS) collects data regarding reimbursements for services provided to beneficiaries. The files contain everything that a bill was generated for including inpatient (hospital), outpatient (skilled nursing facility, hospice, home health), and noninstitution (physician) services for nearly 98% of the United States population 65 years old and over. Patient demographic information is collected. The data are contained within separate files (Standard Analytical Files (SAF), Medicare Provider and Analysis Review File (Med PAR), Physician/Supplier Part B files), and requires significant statistical prowess to link the files (►Table 1).

Medicaid data are obtained from the Medicaid Statistical Information System, Medicaid Analytic extract, and CMS reports. Information is collected regarding eligibility, service usage, and payment for services of the nearly 60 million low-income and/or disabled adults and children covered by Medicaid.¹⁸

To access CMS Medicare and Medicaid files, a data use agreement (DUA) is required through the Research Data Assistance Center (ResDAC).¹⁹ This process can be lengthy and may take several months. The cost of the data is dependent on the amount requested, ranging from several hundred

Table 1 National administrative databases

Registry	Acronym	Variables	Geography	Website
Healthcare Cost and Utilization Project Nationwide Inpatient Sample Kids Inpatient Database Nationwide ED Sample State Inpatient Database State Ambulatory Surgery Database	HCUP NIS KID NEDS SID SASD	Primary/secondary diagnoses Primary/secondary procedures Admission/ discharge status Patient demographics Provider/hospital characteristics Cost, LOS, insurance Inpatient mortality	Nationwide State State	http://www.ahrq.gov/data/hcup
University Health System Consortium	UHC	Diagnoses on admission Inpatient procedures Severity of index score Admission type Mortality, morbidity, LOS, readmission rates, ICU admission, discharge location Cost Provider/hospital characteristics	Nationwide	http://www.uhc.edu
MEDICARE Centers for Medicare and Medicaid Services	CMS	Inpatient, outpatient, skilled nursing facility services Physician services	Nationwide	http://www.resdac.org/
Medicaid Centers for Medicare and Medicaid Services	CMS	Eligibility basis Patient demographics Services provided Prescription drugs		http://www.resdac.org/

Abbreviations: LOS, length of stay; ICU, intensive care unit.

dollars to more than \$10,000. Computer and hardware requirements depend on the quantity of data requested. To overcome some of these difficulties, seminars are offered by ResDac to assist in navigating the complex datasets.

Medicare and Medicaid data files may also be used for study of services received by patients as a group (race, age-related, diagnosis) and among subsets (i.e., adjuvant therapy for rectal cancer for minorities). Stratification of patients based on disease characteristics is *not* available in Medicare (i.e., cannot stratify by stage of disease).

Linkage to clinical registries including the Surveillance, Epidemiology, and End Results (SEER) Program, American Hospital Association (AHA), and the American Medical Association (AMA) Masterfile is possible with Medicare data.

The Healthcare Cost and Utilization Project

The Healthcare Cost and Utilization Project (HCUP) is a family of health care databases developed through a Federal–State–Industry partnership and sponsored by the Agency for Healthcare Research and Quality (AHRQ).²⁰ The HCUP databases bring together the data collection efforts of state data organizations, hospital associations, private data organizations, and the Federal government to create a national information resource of patient-level health care data. The HCUP databases facilitate research on issues including cost and accessibility of health services, practice patterns, and treatment outcomes at regional and national levels.

Individual HCUP databases may be purchased. As of 2011, the specific requirements for HCUP databases include a computer with a DVD drive, a hard drive with 15 GB of space

availability, and ability to unzip a WinZip® file. The cost varies per database. For example, Nationwide Inpatient Sample (NIS) data from 1998–2009 is \$50 for students and \$350 for others and comes on a DVD.

The Nationwide Inpatient Sample

A member of the HCUP family, the NIS is the largest all-payer inpatient care database in the United States, representing a 20% stratified sample of U.S. hospitals and ~8 million annual admissions. Patient discharge information is available by year, allowing for national trends to be evaluated (i.e., number of Hartmann's procedures done over time). Variables in the NIS are listed in ► **Table 1**. As NIS includes only inpatient events, inpatient morbidity and mortality may be studied, but is not comparable to 30-day morbidity and/or mortality available in other databases. Although each unique admission has several diagnoses available to track variables such as comorbidities and demographics, it is often difficult to determine the specifics of an admission that are needed for outcome analysis. For example, a colectomy and *Clostridium difficile* infection could both be coded for a hospital encounter; however, it may not be possible to determine if the colectomy was performed for a *Clostridium* infection, the patient had chronic colonization with *C. difficile*, or if the *Clostridium* infection followed the colectomy. Additionally, no unique patient identifier is assigned in NIS; therefore, readmissions cannot be tracked.

The NIS's large sample size enables analyses of rare conditions, such as congenital anomalies; uncommon treatments, such as organ transplantation; and special patient

populations, such as the uninsured. For most states, the NIS includes hospital identifiers that permit linkages to the American Hospital Association (AHA) Annual Survey Database (Health Forum, LLC © 2010). This linkage permits investigation of both physician and hospital characteristics on patient outcomes including academic affiliation and surgeon/hospital volume.

University Health Consortium

The University Health Consortium (UHC) is an alliance of 116 academic medical centers and 264 of their affiliated hospitals representing ~90% of the nation's nonprofit academic medical centers. It provides patient, hospital, and economic outcomes across different centers. The UHC clinical database resource manager (CDB/RM) includes several variables as demonstrated in ►Table 1.

The UHC developed a severity of illness (SOI) score for risk adjustment of their patients. The SOI score is computed using patient variables including comorbidities, age, and diagnoses. Although it has been validated by AHRQ, it is not used by other databases making comparisons between databases challenging.

Similar to NIS, UHC permits investigation of whether changes in outcomes (morbidity/mortality) or resource utilization (intensive care unit [ICU] admission, length of stay [LOS], hospital cost) are associated with surgeon subspecialty, hospital type, or surgeon/hospital volume.

To access UHC data, an institution must be a member of the consortium. An annual fee related to hospital size is assessed. Data may then be requested and downloaded as a secure file. Specific computer requirements depend on the amount of data requested.

Clinical

Unlike administrative databases, clinical databases were developed with specific clinical goals in mind. Cancer registries track incidence, treatment, and differences in cancer care. Disease-specific information may include stage, tumor characteristics, treatment modalities, and outcomes. Quality improvement databases aim to assess the quality of patient care. Preselected quality measures including intraoperative measures (transfusion requirements, operative time), 30-day morbidity (wound infections, ICU admission, pneumonia) and mortality are reported.

Similar to administrative databases, clinical databases have limitations. Clinical databases may be regional or include areas with large referral centers (i.e., cancer centers), making findings less applicable to the general population.²¹ Research questions are limited to the included variables in the database. Limited data on subjective measures including quality of life and patient satisfaction are available.

The Surveillance, Epidemiology, and End Results Program

The Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute (NCI) collects information on cancer incidence and survival in the United States.

Currently, it collects and publishes cancer incidence and survival data from 17 population-based cancer registries covering ~28% of the U.S. population.²² Included variables are demonstrated in ►Table 2 and include many cancer-related variables. However, although SEER provides detailed information about cancer stage and treatment at the time of diagnosis, therapy completion and long-term outcomes other than death are not accessible. In addition, the SEER database population is predominantly Medicare/Medicaid based, and it tends to have a bias toward older subjects and among the older records, toward Caucasian subjects. SEER also lacks recurrence, radiation, and chemotherapy data, which prohibit drawing direct inference about these important factors regarding oncological outcomes.

SEER data are available for download after completion of a data use agreement for a nominal fee. No specific computer requirements are required. SEER provides a variety of resources to help investigators including SEER*Stat a statistical software program to facilitate analysis of SEER data.²³

National Cancer Database

The National Cancer Database (NCDB) is a joint program of the Commission on Cancer (CoC) of the American College of Surgeons (ACoS) and the American Cancer Society (ACS). It is a nationwide oncology outcomes database for more than 1,500 commission-accredited cancer programs in the United States and Puerto Rico. Data on nearly all types of cancer are tracked using a standardized form including patient characteristics, cancer staging, tumor histological characteristics, treatment, and outcomes (survival). The NCDB captures ~75% of new cancer diagnoses. The main limitation is that cohorts are not population-based, but rather identified from hospitals where they present for diagnosis and/or treatment, thereby limiting the generalizability of the patient population and management algorithms.

The NCDB has web-based tools to evaluate and compare cancer care given by individual institutions with state, regional, and national levels. Data are accessible at no cost to members of the American College of Surgeons after an application process and there are no specific computer requirements.

Cancer Care Outcomes Research and Surveillance Consortium

The NCI established the Cancer Care Outcomes Research and Surveillance (CanCORS) Consortium to evaluate and improve the care and outcomes of patients with lung and colorectal cancer. This consortium is limited to five geographically defined regions, five integrated health delivery care systems, and 15 Veterans Health Administration (VA) hospitals representing ~10% of the U.S. population. Information is extracted from medical records, telephone interviews, and surveys. This database includes specific patient regarding lung and colorectal cancer and also includes socioeconomic, race, and educational status. Outcomes, including patient experience and rating of care were collected by surveys postdiagnosis (4 and 12 months). While providing information on patient perceptions of care, inherent limitations include attrition and

Table 2 National clinical databases

Registry	Acronym	Variables	Geography	Website
The Surveillance, Epidemiology, and End Results Program National Cancer Institute	SEER NCI	Stage/date of diagnosis Primary disease site, therapy Mortality Demographics	17 cancer registries covering ~28% population	http://seer.cancer.gov/
National Cancer Database	NCDB	Patient/hospital characteristics Stage, tumor histology, treatment 6 secondary diagnoses	1450 hospitals	http://www.facs.org/cancer/ncdb/
Cancer Care Outcomes and Research Consortium	CanCORS	ICD oncology codes 6 secondary diagnoses Mortality, stage, comorbidities	5 regions 5 health care systems 15 VA hospitals	http://outcomes.cancer.gov/cancors/
National Surgical Quality Improvement Program American College of Surgeons Veterans Affairs	NSQIP NSQIPACS NSQIPVA	Preoperative risk factors Intraoperative data, Patient demographics Outcomes Procedures 30-Day morbidity/mortality	Participating hospitals nationwide	http://site.acsnsqip.org/
Automated Central Tumor Registry U.S. Department of Defense	ACTUR	Date of diagnosis, date of death Stage, tumor grade Patient demographics	U.S. Department of Defense	Available on request
National Trauma Data Bank	NTDB	Patient demographics Injuries Hospital demographics	National sample Level I/II trauma centers	http://www.facs.org/trauma/ntdb/index.html

recall bias. However, on the positive side, this level of patient information is not available in other large databases. Access to the data requires submission of a proposal that must be reviewed and approved by the CanCORS panel. No specific computer capabilities are required.

National Surgical Quality Improvement Program

The National Surgical Quality Improvement Program (NSQIP®) was initially developed in the VA health care system and has now become a nationally validated, risk-adjusted, outcomes-based program. The aim of NSQIP is to measure and improve the quality of surgical care across surgical specialties. The American College of Surgeons developed ACS NSQIP® to provide a prospective, peer-controlled, validated database of preoperative variables to 30-day postoperative surgical outcomes based on clinical data for patients older than 15 years of age. Trained surgical clinical reviewers collect the data with routine audits to assess accuracy.

Strengths of NSQIP include the wealth of patient clinical data including preoperative comorbidities, operating room information (estimated blood loss, transfusions, time), and 30-day morbidity and mortality. Additional information unique to NSQIP includes postgraduate year of the resident involved with the surgery.

Similar to the other large databases, it often lacks details specific to a patient's hospital course, details of specific medical comorbidities and medications (i.e., immunosuppression use), and indications for surgery (i.e., bleeding, refractory medical management, or toxic megacolon). Patient data are limited to

30-days postdischarge; therefore, readmissions, complications, and deaths after this period are not available. Not all patients are included, but rather a select sampling of patients. To comply with the Health Insurance Portability and Accountability Act (HIPAA), all dates are removed. Most notable for surgical outcomes research is the date of surgery, although additional variables including days to certain events are included (i.e., superficial wound infection, pneumonia). Access to NSQIP data is quite expensive. Annual fees may exceed \$100,000 per year, as a full-time research nurse is required for data extraction. NSQIP has now defined several options to make it more affordable for hospital participation (Essentials, Small & Rural, Pediatric). Individual cost and computer requirements vary with the option selected.

National Trauma Data Bank

The National Trauma Data Bank (NTDB) research dataset (RDS) contains all records submitted to the NTDB for specified admission years. The NTDB national sample program (NSP) is a national probability sample of 100 Level I and II trauma centers in the United States. Data points include admission/discharge status, patient demographics, injury and diagnosis (ICD-9), procedure codes, injury severity, and outcome variables.

Data may be purchased by year or aggregate and cost ranges from \$200 to \$5,000 depending on data selected and whether the purchaser is from a not-for-profit group or a commercial group. Computer requirements depend on data acquired.

Automated Central Tumor Registry

The Automated Central Tumor Registry (ACTUR) was established in 1986 to meet the American College of Surgeons (ACS) requirements for a comprehensive cancer data reporting system and became the cancer database and clinical tracking system for the Department of Defense (DoD). Military medical treatment facilities (MTFs) report cancer data on all DoD beneficiaries, including active-duty military personnel and their family members, retired military personnel, and Reserve and National Guard personnel. Patient demographics, tumor stage and characteristics, geographic information, and outcomes are recorded. Overall survival may be assessed, but disease-specific survival is not available as cause of death is not recorded. Access to the ACTUR databases requires application through the U.S. Department of Defense. The military health care system cohort represented by this database is considerably younger than the general U.S. population, though remains ethnically diverse, and is unique in that its care is based on an equal-access-to-care mandate across the spectrum of diagnosis, treatment, and follow-up.

Linked Databases

SEER-Medicare

The SEER-Medicare database links the CMS Medicare database to the SEER cancer registry. Data are provided for two cohorts; patients with cancer and a random 5% sampling of Medicare beneficiaries residing in the SEER areas who do not have cancer. SEER data including cancer incidence, site, stage, initial treatment, and vital status is linked with Medicare claims for hospital stays, physician and laboratory services, hospital outpatient claims, and home health/hospice bills. Census tract and zip code data are available and can be used to extrapolate patient socioeconomic data. Physician and hospital characteristics include bedsize, hospital ownership, and physician specialty.

The linked SEER-Medicare database allows for study of cancer care, as the data are longitudinal, preexisting conditions may be ascertained, and treatment and outcomes may be evaluated. Noncovered services are not included in the database including prescription drugs and long-term care.

Linking the SEER-Medicare files requires an experienced data analyst, as the files are large and complex. Seminars are held several times a year to introduce programmers and investigators to SEER-Medicare and include database set-up and data analysis.

State Databases

Many state-based databases are available for research purposes. Each database has defined patient demographics, clinical data, geographic information, and hospital data. These vary state to state and can be further investigated based on the state of interest.

Analyses

Once a hypothesis has been generated and a database selected, data analysis may begin. Experience in epidemi-

ology and biostatistics is necessary to code the data, run the analysis, and understand the data limitations. Academic institutions have begun to consolidate resources and assemble research groups comprised of clinical investigators, biostatisticians, epidemiologists, data analysts, and research assistants focused on outcomes research. With institutional support, these research teams provide the depth of knowledge to investigate outcomes and critically analyze data. If this research support is unavailable, identification of an experienced mentor and statistical assistance is critical.

There are many statistical software programs available for data analysis depending on the researchers' experience and training. Commonly used programs include Statistical Analysis System (SAS), Statistical Package for the Social Sciences (SPSS), and Stata Data Analysis and Statistical Software (STATA). These programs can be complex and may require additional server space and computer-processing ability depending on the size of the database. For novice investigators, courses are available for training in both programming and statistical analysis.

Conclusion

Outcomes research using large volume databases are increasingly utilized to investigate surgical outcomes and evaluate quality outcome measures. These administrative and clinical databases serve as well-powered tools to evaluate regional treatment variation and disparities in care. Implementation of quality improvement databases including NSQIP have been shown to decrease morbidity and mortality in participating institutions²⁴ while concurrently providing reliable data for clinical trials. Use of population-based data permits investigations with adequate power that would not be possible with more limited single-site clinical data. Careful study design, appropriate database selection, and rigorous analyses allow investigators to answer key clinical questions in surgery.

References

- 1 Rhodes RS, Biester TW. Certification and maintenance of certification in surgery. *Surg Clin North Am* 2007;87(4):825–836, vi
- 2 Birkmeyer JD, Shahian DM, Dimick JB, et al. Blueprint for a new American College of Surgeons: National Surgical Quality Improvement Program. *J Am Coll Surg* 2008;207(5):777–782
- 3 Birkmeyer NJ, Birkmeyer JD. Strategies for improving surgical quality—should payers reward excellence or effort? *N Engl J Med* 2006;354(8):864–870
- 4 Morris AM, Wei Y, Birkmeyer NJ, Birkmeyer JD. Racial disparities in late survival after rectal cancer surgery. *J Am Coll Surg* 2006;203(6):787–794
- 5 Purves H, Pietrobon R, Hervey S, Guller U, Miller W, Ludwig K. Relationship between surgeon caseload and sphincter preservation in patients with rectal cancer. *Dis Colon Rectum* 2005;48(2):195–202, discussion 202–204
- 6 Deyo RA, Taylor VM, Diehr P, et al. Analysis of automated administrative and survey databases to study patterns and outcomes of care. *Spine (Phila Pa 1976)* 1994;19(18, Suppl):2083S–2091S

- 7 Abdullah F, Ortega G, Islam S, et al. Outcomes research in pediatric surgery. Part 1: overview and resources. *J Pediatr Surg* 2011;46(1):221–225
- 8 Greenblatt DY, Rajamanickam V, Pugely AJ, Heise CP, Foley EF, Kennedy GD. Short-term outcomes after laparoscopic-assisted proctectomy for rectal cancer: results from the ACS NSQIP. *J Am Coll Surg* 2011;212(5):844–854
- 9 Schrag D, Panageas KS, Riedel E, et al. Surgeon volume compared to hospital volume as a predictor of outcome following primary colon cancer resection. *J Surg Oncol* 2003;83(2):68–78, discussion 78–79
- 10 Begg CB, Cramer LD, Hoskins WJ, Brennan MF. Impact of hospital volume on operative mortality for major cancer surgery. *JAMA* 1998;280(20):1747–1751
- 11 Motheral BR, Fairman KA. The use of claims databases for outcomes research: rationale, challenges, and strategies. *Clin Ther* 1997;19(2):346–366
- 12 Cohen ME, Bilimoria KY, Ko CY, Hall BL. Development of an American College of Surgeons National Surgery Quality Improvement Program: morbidity and mortality risk calculator for colorectal surgery. *J Am Coll Surg* 2009;208(6):1009–1016
- 13 Chang DC, Talamini MA. A review for clinical outcomes research: hypothesis generation, data strategy, and hypothesis-driven statistical analysis. *Surg Endosc* 2011;25(7):2254–2260
- 14 Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical information systems. Implications for outcomes research. *Ann Intern Med* 1993;119(8):844–850
- 15 Cooper GS, Yuan Z, Stange KC, Dennis LK, Amini SB, Rimm AA. Agreement of Medicare claims and tumor registry data for assessment of cancer-related treatment. *Med Care* 2000;38(4):411–421
- 16 Best WR, Khuri SF, Phelan M, et al. Identifying patient preoperative risk factors and postoperative adverse events in administrative databases: results from the Department of Veterans Affairs National Surgical Quality Improvement Program. *J Am Coll Surg* 2002;194(3):257–266
- 17 Fisher ES, Baron JA, Malenka DJ, Barrett J, Bubolz TA. Overcoming potential pitfalls in the use of Medicare data for epidemiologic research. *Am J Public Health* 1990;80(12):1487–1490
- 18 Medicaid Available at: <http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Population/By-Population.html>. Accessed May 29, 2012
- 19 Research Data Assistance Center (ResDAC) Available at: <http://www.resdac.org>. Accessed May 29, 2012
- 20 Healthcare Cost and Utilization Project (HCUP) Overview of HCUP: Available at: <http://www.hcup-us.ahrq.gov/overview.jsp>. Accessed May 29, 2012
- 21 Weiss JM, Pfau PR, O'Connor ES, et al. Mortality by stage for right-versus left-sided colon cancer: analysis of surveillance, epidemiology, and end results—Medicare data. *J Clin Oncol* 2011;29(33):4401–4409
- 22 Surveillance Epidemiology and End Results (SEER) Available at: <http://seer.cancer.gov/about/overview.html>. Accessed May 30, 2012
- 23 SEER*Stat Available at: <http://seer.cancer.gov/seerstat/>. Accessed May 29, 2012
- 24 Ozhathil DK, Li Y, Smith JK, et al. Colectomy performance improvement within NSQIP 2005–2008. *J Surg Res* 2011;171(1):e9–e13