
STANDARDS, PRACTICES, AND METHODS FOR THE USE OF ADMINISTRATIVE CLAIMS DATA

John Robst, University of South Florida
Roger Boothroyd, University of South Florida
Paul Stiles, University of South Florida

ABSTRACT

Administrative data are attractive sources of information in research and evaluation studies for numerous reasons including relatively low cost, and the availability of longitudinal information and large subject pools. While many professional organizations set standards for members, there exists a patchwork of practices for researchers to follow when performing research. The purpose of this paper is to outline standards and practices for researchers, and to discuss common analysis issues related to the proper use of administrative data. The discussion focuses on data from the two largest United States government-funded health care programs, Medicare and Medicaid. This focus is chosen due to the wide use of such data, and the sensitive nature of healthcare information.

INTRODUCTION

In a recent paper, Safran et al., (2007) discuss the increasing secondary use of health data for research and other purposes. The authors note that the “lack of coherent policies and standard good practices for secondary use of health data impedes efforts to transform the U.S. health care system” (p. 1). This paper seeks to contribute to this important discussion in two ways. First, a set of standards and practices for researchers to follow is proposed for the acquisition and proper use of administrative data. Second, the literature is reviewed that relates to specific shortcomings with administrative databases and methods to address the problems. The paper is geared towards students with an interest in health economics, but may also be useful to other students and established researchers given the increasing use of administrative data (both health-related and otherwise). The goal is to help

researchers use administrative data correctly so that policy makers can have greater confidence in findings, and consequently research can have a greater effect on public policy.

Public health care programs in the U.S. such as Medicare and Medicaid finance health care for millions of people. The information collected as a result of health care delivery, enrolling members, and reimbursing for services is referred to as administrative data (Iezzoni, 1997). Despite widespread use for research purposes, there exist limited standards and practices for researchers to adhere to in using administrative data (Retchin & Ballard, 1998; Safran, et al., 2007). In addition, while undergraduate and graduate students in economics (and other social sciences) encounter a wide array of courses during their education, few academic programs teach students how to acquire and properly use data.

This paper focuses on data from the two largest government-funded health care programs, Medicare and Medicaid, but the issues discussed in this paper apply to all types of administrative records. The focus was chosen because of the sensitive nature and yet widespread use of such data, the increased vulnerability of the subjects, and the evolving U.S. federal regulatory landscape for healthcare information in general. Examples are discussed based on experiences during the lead author's five years at the Centers for Medicare & Medicaid Services (CMS), the government agency that oversees the programs.

ADVANTAGES AND DISADVANTAGES TO ADMINISTRATIVE DATA

First, let's review a few of the advantages and shortcomings of using administrative data for research. There are a number of advantages to administrative data (Iezzoni, 2002; Pandiani & Banks, 2003; Roos, Menec, & Currie, 2004; Roos et al., 2008). It is conceivable to study (almost) all individuals age 65 and above with Medicare enrollment and claims data. The use of population based data enables questions to be considered that could not be addressed with a sample. However, due to cost considerations and the sheer size of the databases, almost all studies use a sample. For example, as discussed in more detail later in the paper, much research uses a 5% sample of Medicare beneficiaries which is approximately 800,000 people. Despite being a small proportion of beneficiaries, the sample size remains substantial and limits concerns about the generalizability of results found in small sample studies. In addition, the large size also allows for adequate numbers of minorities for statistical analysis.

The records are not limited to specific types of setting (e.g., hospitals). Information can be longitudinal covering individuals and institutions across many years. Confidentiality can be maintained due to the large sample sizes. The data exist, and thus are relatively inexpensive to acquire compared to primary data collection, plus the low cost also allows for easy replication of previous studies. Survey attrition due to a loss of contact or refusal to participate is also minimized.

There are, however, many potential problems with the use of administrative data (Retchin & Ballard, 1998; Drake & McHugo, 2003). Such problems include a lack of information on the reliability or accuracy. Public use files may not be available for several years, reducing usefulness for current policy questions. Large samples can lead to statistically significant results that are not very meaningful, as even very small effects are precisely measured. Similarly, researchers may look for questions to fit the data, rather than forming questions and then looking for the appropriate data. Medicaid and Medicare enrollment and claims records include protected health information under the Health Insurance Portability and Accountability Act (HIPAA) and therefore require stringent privacy protection measures.

Due to such potential problems, users should adhere to standards on data use. While most professional organizations establish standards for members, there are no clear standards and practices for users of administrative data to follow. However, appropriate use is crucial in order to increase public confidence in the use of sensitive health care information for research purposes, and for federal agencies to continue to allow access to the data (Safran, et al., 2007).

THE RESEARCH PROTOCOL: DATA ADEQUACY AND ACQUISITION

Acquisition of administrative data typically begins with the development of a detailed research protocol. The protocol is assessed by the data owners to determine whether access should be granted to Medicaid or Medicare enrollment and claims records. A useful resource for researchers developing a protocol, although involving data for Canada, is provided at The Manitoba Centre for Health Policy (MCHP) web site (<http://www.umanitoba.ca/centres/mchp/>). Some of the information is specific to the MHCP mission and data on Manitoba residents, but much of the information applies to administrative data in general.

The protocol should detail the research questions and explain why they are important to the mission of the Medicare and/or Medicaid programs. Given the

inherent concern in releasing sensitive information, research questions need to be of sufficient interest to the data owners to warrant release of the data. The protocol must also identify the specific dataset(s) and justify that the source is appropriate for the proposed analysis. van Eijk, Krist, Avorn, Porsius, & de Boer (2001) created a checklist guidelines to determine whether available data are adequate to answer the research questions. Important considerations used to decide whether the available data are adequate to meet research needs include sample size, whether the claims contain sufficient detail for the study (e.g., diagnoses, procedures, drug and dosing information), accuracy, continuity of variables over time, the ability to link databases, and adequate security and accessibility. In the following sections, several of these considerations are discussed as well as others as they relate to secondary use of health data.

Approvals

An important early step is to understand the process for data acquisition. Most data available from CMS are acquired through the Research Data Assistance Center (ResDAC), a CMS contractor that provides assistance to researchers using Medicare enrollment and claims records. Their web site www.resdac.umn.edu) contains much information on the process for acquisition and the associated cost but provides limited guidance on the proper use of the data. Together, CMS and ResDAC act as gatekeepers and determine who gains access to CMS data. ResDAC and CMS also make available national Medicaid data, referred to as the Medicaid Extract (MAX) files. The MAX files are a combination of the Medicaid enrollment and claims data compiled by each state. Some states make Medicaid data from their state available to researchers, some do not. If you wish to use Medicaid data from a specific state, contact the Medicaid authority and determine whether the data are available and what their process for data acquisition entails.

Consult with data owners

Users should consult with data owners to understand what the data represent and ensure the proposed questions can be appropriately answered with the data. For Medicare data, this may involve discussions with ResDAC personnel and also individuals at CMS who work with the data. There are several reasons for users to seek such consultation. Administrative data are usually compiled for a specific purpose, often related to payment or program monitoring and evaluation. Users

need to understand why the administrative database was created. The reason(s) for collecting the data can have an important impact on the universe covered, data elements, variable definitions, frequency and timeliness, quality, and stability over time. A lack of understanding of what the data represent and how it may be used has lead researchers to propose research questions for which the data are poorly suited (Medi-Cal Policy Institute, 2001).

In addition, given that administrative data are often compiled for internal use by the data owners, documentation is often scant compared to survey data primarily produced for research purposes. Even with proper documentation, owners are a valuable resource for understanding technical details and should be consulted by users. The data owners have knowledge of the issues involved in working with the files, problems with specific variables, are aware of other issues not apparent from reading documentation or examining the data, and can verify that the project design is appropriate.

Such discussions also provide opportunities to clarify variable definitions. For example, Medicare enrollment files note when Medicare is a secondary payer. This occurs primarily when the beneficiary has health insurance coverage through a spouse. The person is labeled as *working aged* despite the fact that the beneficiary is not employed. Consequently, users should not assume the variable name necessarily describes the variable clearly.

Data quality

Data users should always consider the likely quality of the data for the proposed research questions. The accuracy of data is extremely important, particularly for analyses to inform public policy (Robinson & Tataryn, 1997). While the available quantity of information is often large, the accuracy and completeness is sometimes questioned. The Medi-Cal Policy Institute (2001) reported that California's Medicaid managed care data system could not be "used to make sound policy decisions" because data were inaccurate and incomplete. Most administrative data rely on reporting by individuals or firms and the information respondents provide can cause gains or losses to individuals or businesses (Wolf & Helminiak, 1998). In other cases, information can be underreported if unrelated to the gains or losses of individuals or businesses. As such, there may be biases in the information supplied.

Even if the overall database is considered complete and accurate, specific variables may differ in accuracy. Administrative files used to make payment often

have fields that are checked for completeness and reasonableness. As such, these fields are relatively accurate. Other variables may not be checked or edited, especially those that do not affect payment. Users should learn the editing rules used by the owners. Users should determine the likely extent of measurement error and decide whether it should be addressed in the research plan.

The sample

One potential benefit of administrative data is the ability to perform population based research. In theory, Medicare data may be available for all individuals age 65 and above. The analysis of population based data avoids many of the concerns with analyzing samples, whether small or large. All statistics are *actual* statistics, not *sample* statistics. Thus, conclusions can be drawn without concerns about type I or type II errors.

In practice, the Medicare program does not cover everyone age 65 and above. Individuals must qualify for Medicare based on work history (either their own or a spouse's). Some individuals never establish a sufficient work history to qualify for Medicare. For example, certain immigrant groups are less likely to qualify for Medicare because work histories were not established with the Social Security Administration. Thus, even with a database as large as the Medicare enrollment and claims data, users must be aware of who may not be adequately represented in the data and potential biases this may introduce. In addition, given the size of some administrative databases, users should consider whether they have sufficient resources (both computer and financial) to acquire, store, and analyze the data. For example, there are over a billion Medicare claims in a single year.

In almost all cases, researchers use a sample. A five or ten percent sample from a very large database is sufficient for the majority of studies. For example, many researchers use the CMS 5% Medicare Standard Analytical Files (SAF). The standard analytical files contain all enrollment and claims data for 5% of Medicare beneficiaries (approximately 800,000 people) and are created annually by CMS. Because these files are used by many researchers, the cost of acquiring the data is lower than if a researcher requests a special data pull. The SAFs are created by selecting all enrollment and claims records for individuals with 05, 20, 45, 70 or 95 in positions 8 and 9 of the health insurance claim number (i.e., the last two digits of the Medicare identification number). The sample selection criteria for the SAFs allows for individuals to be followed over time, which would not be possible with a true random sample. At the same time, this could be problematic is the last two

digits of the Medicare IDs differed across individuals in a systematic nature. However, the Medicare ID is typically the person's Social Security number (plus characters in the 10th and 11th places to denote the reason for eligibility). The last two digits of a Social Security Number are not systematically assigned based on characteristics of the individual, and thus the SAFs are generally considered to be equivalent to a random sample. If, for example, the sample was pulled based on the first three digits (where are assigned based on geographic location), then the sample would be geographically biased and not representative of the population.

While generally not a concern with large administrative databases, users must consider if the expected number of observations is sufficient to generate meaningful results. In general, power tests should be performed to determine the sample size necessary to have reasonable confidence that statistically significant results can be detected. This step is particularly important if studying a rare disease or treatment. At the same time, given the typical large sample size, users have to interpret the economic significance of their results and not simply rely on statistical significance (Drake & McHugo, 2003).

Researchers must also know the decision rules used to pull the data. For example, studies interested in the frequency of services should know if claims are "final action", or if they include denials, interim bills, or adjustments. The inclusion of interim bills and adjustments will lead to an over count of service frequency, and thus should be excluded during the analysis.

Diagnostic accuracy

Research questions often focus on specific subgroups of individuals with specific diagnoses (e.g., asthma or diabetes). Claims data contain codes that identify specific diseases using the International Classification of Diseases (ICD). ICD codes are five digit codes that can be used to identify individuals with a broad class of diseases or a very specific disease. The first three digits tend to identify a general class (e.g., 250 for diabetes), with the fourth and fifth digits being more specific (e.g., 250.41 denotes type I diabetes with renal manifestations).

Among the issues to consider is whether two years or more of data should be used to identify cases. Dombkowski, Wasilevich, and Lyon-Callo (2005) found that a diagnosis of a chronic disease (asthma) was not observed in every year. Thus, selection of cases based on diagnoses in a single year would undercount the prevalence of a disease. People still have the disease but it did not show up in the

claims data during a year for some reason. Consequently, the identification of individuals with chronic diseases may require multiple years of data.

In addition to diagnoses, prescription drug use might also identify people (e.g., insulin or perhaps metformin use for diabetes). Gilmer et al. (2001) find that the use of prescription drug records substantially increases the estimated prevalence of specific diseases. Caution must be used though since many medications are used to treat multiple conditions, and thus might not indicate a specific disease.

On the other hand, consideration might also be given to whether an individual should be included only if there are at least two records with the diagnosis of interest to rule out incorrect or miscoded chronic diagnoses. The presence of a consistent diagnosis over time provides evidence that the diagnosis is correct. Such concerns arise from studies that compare diagnoses in medical charts and claims. For example, Schwartz et al., (1980) find a relatively poor match between medical charts and claims for Medicaid enrollees; 29% of chart diagnoses of private practitioners, 37% of chart diagnoses in the free standing outpatient clinics; and 46% of diagnoses from outpatient clinics of general hospitals do not match with Medicaid claims. Interested readers should look at Virnig & McBean (2001) for a more thorough discussion of studies that assess reliability by comparing diagnostic data located in charts to claims in the database.

Security

Researchers are responsible for data security, and should have a plan for ensuring that the files cannot be accessed by unauthorized users. Some obvious steps include using automatic screen savers that can only be turned-off with a password if the data reside on an office or personal computer. If storage is on a network, only authorized users should have access, and the data should be behind a firewall if the network is connected to the internet. Email is not a safe way to transmit individually identifiable information unless adequate encryption is used. In addition, user responsibility for the data does not end when the project ends. The data use agreement (DUA) typically specifies whether the data have to be returned to the agency or destroyed.

THE PROTOCOL – DATA ANALYSIS

The protocol must also detail the analysis plan. This section provides an overview of some common analysis issues related to using administrative data.

Such analytical issues include the need to empirically assess quality, differentiate between time trends and program effects, and use medical encounters to account for the differing health status of treatment and comparison groups (Ray, 1997). Much depends on the study questions and design for the specific project. The proposed analysis should meet the standards for institutional review boards and peer reviewed publications.

Studies are discussed below that relate to the analysis issues and the solutions employed by researchers. The studies are not an exhaustive overview of questions that can be analyzed with administrative data. Readers interested in a broader discussion of health care topics that can be addressed with administrative data should see a paper by Roos, Menec, & Currie (2004), and for a broader discussion of how administrative data can be used to answer an array of social research questions see Roos et al., (2008).

Linking records

Users will often need to merge several different data files. Examples of such linkages include combining records from inpatient, outpatient, and physician claims, supplementing claims data with survey data such as the Medicare Current Beneficiary Survey, or matching individuals across years. Privacy concerns may arise when administrative records are linked to other sources and researchers should verify that the data use agreement allows such linkage (Clark, 2004).

Linking may be based on shared identifiers, deterministic matching, or probabilistic matching (Victor & Mera, 2001; Clark, 2004; Roos et al., 2008). Matching records by shared identifiers occurs when there are the same identifiers in data sets (e.g., Social Security Number or Health Insurance Claim number). Most data available from CMS can be matched using individual identifiers. However, researchers may also encounter situations when unique individual identifiers are not available. Deterministic matching examines a subset of variables and matches records that agree on this subset (e.g., name, date of birth, sex). Individuals can have the same name or date of birth or sex, but it is far less likely that different individual in two datasets will have the same name and date of birth and sex. Probabilistic linking matches based on the probability that records refer to the same person. Matching with individual identifiers or deterministic matching is typically used when attempting to draw conclusions about individuals. Probabilistic matching is used when there is limited information on which to base matches (e.g., name, date

of birth, sex). Given the difficulty in precisely matching individuals, probabilistic matching is more appropriate when drawing conclusions about populations.

The use of probabilistic matching is illustrated by Banks & Pandiani (1998). The authors derive estimates of the number of people receiving psychiatric care in state hospitals and general medical settings. Typically, the data sets would be merged based on individual identifiers or deterministic matching to avoid double counting patients who receive care in both sectors. Banks and Pandiani use probabilistic matching based only on gender and birth date to derive estimates of sample overlap, and as a result are able to estimate the number of people receiving psychiatric care. The use of probabilistic matching is likely to increase as concerns with patient privacy lead data owners to restrict the release of information that enables direct or deterministic matching to other data sources.

When records from more than one administrative source are combined it is important to be aware of potential differences in concepts, definitions, reference dates, coverage, and quality. For example, recent attention has focused on merging Medicare and Medicaid claims to study dual eligible beneficiaries (e.g., Liu, Wissoker, & Swett, 2007; Yip, Nishita, Crimmons, & Wilber, 2007). These data originate from different sources that may use different definitions, definitely have different coverage issues, and may have differences in quality. While one might expect data within the Medicare program to have consistent standards, even this may not necessarily be the case. For example, the quality of inpatient hospital claims is generally considered better than physician claims (Retchin & Ballard, 1998).

Empirically assess data quality

While data quality should be assessed for expected accuracy prior to acquisition, quality should also be assessed empirically. Once the data are linked and the sample constructed, users should examine descriptive statistics. Users should check the results for reasonableness, and if possible, compare results with alternative data sources or prior research and attempt to explain differences. Many studies have been published using Medicare and Medicaid data providing researchers with a substantial literature for comparison.

Assessing quality is particularly important when data are hand entered because data errors may be more prevalent. An example of such data entry errors occurs with beneficiary location (SSA state and county codes) in Medicare claims. Research often looks at Medicare utilization across counties in the United States. Analyzing claims, there are approximately 5,000 SSA state/county codes that appear

in the data, far greater than the 3,100 actual counties in the US. What accounts for the erroneous counties? State and county codes are often hand entered, there is no payment issue involved (payments are based on provider location, not beneficiary residence), and the field is not checked for accuracy. Such miscoding may be important for sparsely populated counties where a few miscoded observations can make a difference to the results.

Two approaches are used in the literature to address potential problems with examining geographic variation in prevalence rates across counties. For example, Cooper, Yuan, Jethva, & Rimm (2002) examine county level variation in breast cancer rates using Medicare data. The authors attempt to confirm their findings by comparing prevalence rates to the National Cancer Institute's cancer tracking database (the Surveillance, Epidemiology, and End Results, SEER, program) which tracks approximately 10-15 percent of the U.S. population. While a valid test for large counties, comparing prevalence rates in small counties could be problematic due to the (relatively) small sample size of the SEER database. Holcomb & Lin (2006) examine geographic variation of macular disease in Kansas. Because of the potential for unstable prevalence rates in small counties, the authors aggregated sparsely populated counties into larger geographic units.

Researchers should document their findings regarding quality to enable other researchers to understand why certain observations or variables were included or excluded based on data quality considerations. Documentation can also help researchers compare and reconcile studies so that others understand why decisions were made and potential implications of those choices. These are basic steps that all researchers should perform, but studies are often unable to replicate research because such steps are not taken (Dewald, Thursby, & Anderson, 1986). New users of a data set should review the literature to see how others have handled problems with the data.

Time series analysis

Over time, the Medicare and Medicaid programs have moved toward managed care, case management, and provision of prescription drugs. Consequently, it is increasingly important to track people over time to determine how participation in case management or the provision of certain prescription drugs affects health over time. For example, there has been much discussion about creating a database to track outcomes from prescription drug use after the well documented problems with Vioxx (e.g., Lohr, 2007). Multiple years of the CMS Standard

Analytical Files are often linked to examine changes over time. This is possible because, as discussed earlier, the 5% sample contains all enrollees with HIC numbers that end in specific digits. Thus, with some exceptions (some people die and new enrollees enter the data) the sample contains the same people over time.

A substantial literature using time series analysis considers the changing prevalence of specific diseases over time. For example, Lakshminarayan, Solid, Collins, Anderson, & Herzog (2006) find an increasing prevalence of atrial fibrillation diagnoses between 1992 and 2002, while Salm, Belsky, & Sloan (2006) find an increasing prevalence of eye diseases between 1991 and 2000. Lakshminarayan (2006) partly address diagnostic quality by requiring at least one inpatient claim or two outpatient claims with an atrial fibrillation diagnosis. However, both studies may be overstating the increasing prevalence of such diseases. Physicians were required to report diagnostic data on Medicare claims beginning in the early 1990s. Physician payments are typically based on procedures not diagnoses, and diagnosis is often not necessary to justify a procedure. Over time physicians have reported more thorough diagnostic data. Indeed, diagnostic reporting continues to improve more than a decade later as physicians implement electronic medical records. The point is that if one examines time trends in the prevalence of a disease, one needs to be cautious in looking at diagnostic trends in Medicare claims data. Simply looking at the increased reporting of a diagnosis is likely to overstate the increasing prevalence of a disease. While Salm (2006) at least note this possibility, neither study attempted to account for this in their analysis.

When records from different time periods are linked, they are a very rich source of information for researchers. However, users should understand whether the data will be consistent across time, and why changes may occur. The reasons for collecting the information may change over time, or variable definitions may change, or reporting may have changed. Coverage changes occur on a regular basis in Medicare based on CMS decisions and Congressional mandates. Such changes can have a substantial effect on services provided.

Accounting for individual heterogeneity

Perhaps the biggest challenges in using administrative data are to create a comparison group and decide on the appropriate analytical techniques. In clinical research, randomized control trials allow researchers to assign individuals to treatment and control groups in a random manner. Administrative data do not typically allow for this type of assignment and there are often non-random

differences between individuals that choose a treatment versus no treatment (or an alternative treatment). Pre-treatment differences may bias (typically referred to as sample selection bias) the results if such differences also correlate with the outcome.

Selection issues are common in research using administrative data, requiring researchers to account for differences between individuals. For example, administrative claims are often used to assess quality of care and examine outcomes from patient care. Hospital quality has been considered by many researchers because hospital administrative data are generally considered to be relatively high quality (e.g., Krumholz et al., 2006; Ross et al., 2007). Quality of care by physicians is also considered by Schatz et al., (2005). However, hospitals and physicians that have the most complex cases are more likely to have the highest complication and mortality rates. Consequently, accounting for case-mix is crucial to comparing the care provided across medical care settings or outcomes from alternative treatments.

There are several methods used to account for pre-treatment differences. The first two methods focus on accounting for observed differences between individuals. Many studies use risk adjusted models where control variables thought to be correlated with the outcome and the independent variable of interest (e.g., hospitals, physicians, treatment, gender, race, etc.) are included in a regression specification. Popescu, Vaughan-Sarrazinn, and Rosenthal (2007) examine racial differences in mortality after acute myocardial infarction. The authors control for sociodemographics, comorbidity, and illness severity to account for factors potentially correlated with the outcome (mortality) and variable of interest (race).

A variant on this approach is to use a diagnosis-based risk score as a measure of health (e.g., Ross et al., 2007). The score represents a measure of overall health status based on demographics and diagnoses. CMS and many States use diagnosis-based risk scores to determine compensation for managed care plans (e.g., Pope et al., 2004). While a useful measure, many researchers do not compute the scores correctly. This occurs because the models were developed using diagnoses from specific provider types (e.g., physician specialty). While this detail is contained in the technical instructions for managed care plans to submit data, it is not included in the risk adjustment publications or software published by CMS. Since most researchers do not discuss their research with people who work on risk adjustment at CMS, they often include too much diagnostic data when computing individual risk scores and overstate risk scores. As pointed out earlier, existing documentation may not provide all needed information, but such information can be learned by consulting with knowledgeable individuals. The example suggests

that users should initiate such discussions regardless of whether the researcher is aware of a lack of information.

A potential problem with risk adjusted regression models, regardless of whether specific characteristics of risk are used or an overall risk score, is that the comparison groups may not have the control variables in common. For example, if a treatment group is primarily old and a control group is primarily young, then conclusions regarding the effect of treatment may be biased given linearity assumptions in regression modeling.

Propensity score matching has become a popular alternative to regression methods in social science research for addressing selection issues when analyzing administrative files (Rosenbaum & Rubin, 1983; Imbens, 2000). Matching techniques mimic a random experiment by matching individuals in the treatment and control groups based on observed characteristics. The observed characteristics are used to estimate the probability of receiving treatment. Individuals with similar probabilities of treatment are compared, some who do and some who do not receive the treatment, to determine the effect of treatment. Using the age example, the young people in the treatment and control groups would be matched, while the older individuals in the treatment and control groups would be matched. Outcomes are then appropriately compared for similar individuals.

Numerous articles use propensity score methods to examine treatment effects when using administrative data. For example, Berg & Wadhwa (2007) examine the effects of a disease management program for elderly patients with diabetes. Propensity score methods are used to match observations in the treatment group with people in a control group who did not participate in the disease management program. Similarly, Krupski et al., (2007) examine the effects of receiving androgen deprivation therapy for individuals with prostate cancer on skeletal complications. Individuals receiving therapy are matched to individuals not receiving therapy by age, geographic region, insurance plan, and index year.

There is, however, debate about whether matching actually mimics a random experiment (Agodini & Dynarski, 2004; Smith & Todd, 2005). Research attempting to validate propensity score matching uses experimental data, and attempts to replicate the experimental results by reexamining the data using matching techniques. In other words, the data are examined under the assumption that assignment was not random and may be subject to selection biases. The majority of studies find the results from experimental data and matching methods are not similar. Thus, while matching methods may be useful, they should not be viewed as a perfect solution to problems with sample selection.

One potential problem with each of the above methods is the reliance on observed data. As such, the development of risk scores and propensity scores is challenging with administrative claims that often lack key clinical detail (Iezzoni, 1997). This issue is particularly salient for research on provider quality. Iezzoni (1997) suggests that administrative data be used as a screening tool to highlight areas for further investigation, not to draw conclusions about quality. Information on the process and appropriateness of care may not be adequate to provide accurate measures of provider quality. In general, all studies involve some degree of unobserved data. Instrumental variables methods may be appropriate if unobservable characteristics are thought to be important to the analysis. Of course, it can be extremely challenging to find suitable instruments. In conclusion, controlling for differences between treatment and control groups, or between patients seen at different hospitals, or between any two comparison groups, is crucial to drawing proper conclusions.

Research tools

There are many tools available to researchers on the internet and it may also be useful to utilize publicly available modules to develop important measures. Consistency across studies is increased if users can access such modules. Such publicly available information is typically tested by numerous users and is likely to be accurate. Much research requires manipulation of the data to create the analysis files and measures needed to answer the research questions. The internet allows researchers to utilize publicly available programs and modules that enable accurate creation of health measures such as the Charlson Index. The Manitoba Centre for Health Policy (MCHP) web site provides a web-based repository of useful tools for conducting research using administrative data (Roos, Soodeen, Bond, & Burchill, 2003). Some of the modules apply specifically to data available from the MCHP, but there are a number of statistical tools for analysis that can apply to a variety of administrative claims sources.

CONCLUSION

This paper has outlined some practice guidelines for the use of administrative data. While administrative data have great potential, there are also many pitfalls. Research using secondary data will benefit the health care of Americans only if the data are appropriately used. The growing use of such records

in research and evaluation necessitates that guidelines be developed and discussed such that the conclusions from research are valid. We hope the guidelines presented in this paper generate further discussion of the appropriate use of such data.

In summary, users of administrative data should develop a research protocol that: presents the research questions including a justification of why the research questions are important to the data owners, assesses whether the data are appropriate for the research questions (i.e., quality, sample size, available variables, and ability to link records) through reviews of the literature and discussions with the data owners, details the security plan including where the data will be stored and how access will be controlled, presents the analysis plan including an empirical assessment of the data quality and the statistical techniques that will be used to answer the research questions, discusses how potential data shortcomings will be addressed, and describes steps that will enable replication by other researchers.

Clearly, there is a need for such standards and practices in the use of administrative data given the continued increase in use. Huax (2005) outlines some of the current trends and his views on upcoming changes in health information systems. The trend continues to be towards using administrative data to inform patient care, strategic management, and clinical and epidemiological research. The future is likely to move towards the development of comprehensive electronic medical records that include information from multiple or all payers. As administrative data become more comprehensive and complex, developing and utilizing standards and practices will become even more important in the future.

REFERENCES

- Agodini, R., & Dynarski, M. (2004). Are experiments the only option? A look at dropout prevention programs. *Review of Economics and Statistics*, 86, 180-194.
- Banks, S.M., & Pandiani, J.A. (1998). The use of state and general hospitals for inpatient psychiatric care. *American Journal of Public Health*, 88, 448-451.
- Berg, G.D., & Wadhwa, S. (2007). Health services outcomes for a diabetes disease management program for the elderly. *Disease Management*, 10, 226-234.
- Clark, D.E. (2004). Practical introduction to record linkage for injury research. *Injury Prevention*, 10, 186-191.

-
- Cooper, G.S., Yuan, Z., Jethva, R.N., & Rimm, A.A. (2002). Use of Medicare claims data to measure county-level variation in breast carcinoma incidence and mammography rates. *Cancer Detection and Prevention*, 26, 197-202.
- Dewald, W.G., Thursby, J.G., & Anderson, R.G. (1986). Replication in empirical economics: The *Journal of Money, Credit, and Banking* project. *American Economic Review*, 76, 587-603.
- Dombkowski, K.J., Wasilevich, E.A., & Lyon-Callo, S.K. (2005). Pediatric asthma surveillance using Medicaid claims. *Public Health Reports*, 120, 515-524.
- Drake, R. & McHugo, G. (2003). Large data sets can be dangerous. *Psychiatric Services*, 54, 133.
- Gilmer, T., Kronick, R., Fishman, P. et al. (2001) The Medicaid Rx model: Pharmacy-based risk adjustment for public programs, *Medical Care*, 39, 1188-1202.
- Haux, R. (2005). Health information systems – past, present, future. *International Journal of Medical Informatics*, September 15, 2005.
- Holcomb, C.A., & Lin, M.C. (2005). Geographic variation in the prevalence of macular disease among elderly Medicare beneficiaries in Kansas. *American Journal of Public Health*, 95, 75-77.
- Iezzoni, L.I. (2002). Using administrative data to study persons with disabilities. *The Milbank Quarterly*, 80, 347-378.
- Iezzoni, L.I. (1997). Assessing quality using administrative data. *Annals of Internal Medicine*, 127, 666-674.
- Imbens, G.W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87, 706-710.
- Krumholz, H.M., Wang, Y., Mattera, J.A., Wang, Y.F., Han, L.F., Ingber, M.J., Roman, S., & Normand, S.L.T. (2006). An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation*, 113, 1683-1692.
- Krupski, T.L., Foley, K.A., Baser, O., Long, S., Macarios, D., & Litwin, M.S. (2007). Health care cost associated with prostate cancer, androgen deprivation therapy and bone complications. *Journal of Urology*, 178, 1423-1428.

- Liu, K., Wissoker, D., & Swett, A. (2007). Nursing home use by dual-eligible beneficiaries in the last year of life. *Inquiry*, 44, 88-103.
- Lakshminarayan, K., Solid, C.A., Collins, A.J., Anderson, D.C., & Herzog, C.A. (2006). Atrial fibrillation and stroke in the general Medicare population: A 10 year perspective, 1992-2002. *Stroke*, 37, 1969-1974.
- Lohr, K.N. (2007). Emerging methods in comparative effectiveness and safety: Symposium overview and summary. *Medical Care*, 45, S5-S8.
- Medi-Cal Policy Institute. (2001). *From Provider to Policymaker: The Rocky Path of Medi-Cal Managed Care Data*.
- Pandiani, J. & Banks, S. (2003). Large data sets are powerful. *Psychiatric Services*, 54, 745.
- Pope, G.C., Kautter, J., Ellis, R.P., Ash, A.S., Ayanian, J.Z., Iezzoni, L.I., Ingber, M.J., Levy, J.M., & Robst, J. (2004). Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financing Review*, 25(4), 119-141.
- Popescu, I., Vaughan-Sarrazin, M.S., & Rosenthal, G.E. (2007). Differences in mortality and use of revascularization in black and white patients with acute MI admitted to hospitals with and without revascularization services. *Journal of the American Medical Association*, 297, 2489-2495.
- Ray, W.A. (1997) Policy and program analysis using administrative databases. *Annals of Internal Medicine*, 127, 712-718.
- Retchin, S.M., & Ballard, D.J. (1998). Establishing standards for the utility of administrative claims data. *Health Services Research*, 32, 861-866.
- Robinson, J. & Tataryn, D. (1997). Reliability of the Manitoba mental health management information system for research. *Canadian Journal of Psychiatry*, 42 744-749.
- Roos, L., Borwnell, M., Lix, L., Roos, N., Walld, R., & MacWilliam, L. (2008). From health research to social policy: Privacy, methods, approaches. *Social Science & Medicine*, 66, 117-129.
- Roos, L.L., Menec, V., & Currie, R.J. (2004). Policy analysis in an information-rich environment. *Social Science and Medicine*, 58, 2231-2241.

-
- Roos, L.L., Soodeen, R.A., Bond, R., & Burchill, C. (2003). Working more productively: Tools for administrative data. *Health Services Research*, 38, 1339-1357.
- Ross, J. Cha, S. Epstein, A., Wang, Y., Bradley, E., Herrin, J., Lichtman, J., Normand, S., Masoudi, F., & Krumholz, H. (2007). Quality of care for acute myocardial infarction at urban safety-net hospitals. *Health Affairs*, 26, 238-248.
- Rosenbaum, P.R., & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.
- Safran, C., Bloomrosen, M., Hammond, W.E., Labkoff, S., Markel_fox, S., Tang, P.C., & Detmer, D.E. (2007). Toward a framework for the secondary use of health data: An American Medical Informatics Association white paper. *Journal of the American Medical Informatics Association*, 14, 1-9.
- Salm, M., Belsky, D., & Sloan, F.A. (2006). Trends in cost of major eye diseases to Medicare, 1991-2000. *American Journal of Ophthalmology*, 142, 976-982.
- Schatz, M., Nakahiro, R., Crawford, W., Mendoza, G., Mosen, D., & Stibolt, T.B. (2005). Asthma quality-of-care markers using administrative data. *Chest*, 128, 1968-1973.
- Schwartz, A.H., Perlman, B.B., Paris, M., Schmidt, K., & Thornton, J.C. (1980). Psychiatric diagnoses as reported to Medicaid and as recorded in patient charts. *American Journal of Public Health*, 70, 406-408.
- Smith, J. & Todd, P.E. (2005). Does matching overcome Lalonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125, 305-353.
- van Eijk, M., Krist, L., Avorn, J., Porsius, A., & de Boer, A. (2001). Do the research goal and databases match? A checklist for a systematic approach. *Health Policy*, 58, 263-274.
- Victor, T.W., & Mera, R.M. (2001). Record linkage of health care insurance claims. *Journal of the American Medical Informatics Association*, 8, 281-288.
- Virnig B.A., & McBean A.M. (2001). Using administrative data for public health surveillance and planning. *Annual Review of Public Health*, 22, 213-230.
- Wolf, N. & Helminiak, T.W. (1998). Nonsampling measurement error in administrative data: Implications for economic evaluations. *Health Economics*, 5, 501-512.

Yip, J., Nishita, C.M., Crimmons, E.M., & Wilber, K.H. (2007). High-cost users among dual eligibles in three care settings. *Journal of Health Care for the Poor and Underserved*, 18, 950-965.