Volume 6 • Number 2 • 2003
VALUE IN HEALTH

A Checklist for Retrospective Database Studies—Report of the ISPOR Task Force on Retrospective Databases

Brenda Motheral, MBA, PhD,[1] John Brooks, PhD,[2] Mary Ann Clark, MHA,[3] William H. Crown, PhD,[4] Peter Davey, MD, FRCP,[5] Dave Hutchins, MBA, MHSA,[6] Bradley C. Martin, PharmD, PhD,[7] Paul Stang, PhD[8]

[1]Express Scripts, Maryland Heights, MO, USA; [2]College of Pharmacy, University of Iowa, Iowa City, IA, USA; [3]Boston Scientific Corporation, Natick, MA, USA; [4]The Medstat Group, Cambridge, MA, USA; [5]Department of Clinical Pharmacology, University of Dundee, Dundee, UK; [6]Advanced PCS Health Systems, Inc., Scottsdale, AZ, USA; [7]College of Pharmacy, University of Georgia, Athens, GA, USA; [8]Galt Associates, Inc., Blue Bell, PA, USA

ABSTRACT

Introduction: Health-related retrospective databases, in particular claims databases, continue to be an important data source for outcomes research. However, retrospective databases pose a series of methodological challenges, some of which are unique to this data source.

Methods: In an effort to assist decision makers in evaluating the quality of published studies that use health-related retrospective databases, a checklist was developed that focuses on issues that are unique to database studies or are particularly problematic in database research. This checklist was developed primarily for the commonly used medical claims or encounter-based databases but could potentially be used to assess retrospective studies that employ other types of databases, such as disease registries and national survey data.

Results: Written in the form of 27 questions, the checklist can be used to guide decision makers as they consider the database, the study methodology, and the study conclusions. Checklist questions cover a wide range of issues, including relevance, reliability and validity, data linkages, eligibility determination, research design, treatment effects, sample selection, censoring, variable definitions, resource valuation, statistical analysis, generalizability, and data interpretation.

Conclusions: For many of the questions, key references are provided as a resource for those who want to further examine a particular issue.

Keywords: claims databases, outcomes research, research design, statistics.

Introduction

What Is the Purpose of This Checklist?

This checklist is intended to assist decision makers in evaluating the quality of published studies that use health-related retrospective databases. Numerous databases are available for use by researchers, particularly within the United States. Because the databases have varying purposes, their content can vary dramatically. Accordingly, the unique advantages and disadvantages of a particular database must be borne in mind. In reviewing a database study, it is important to assess whether the database is suitable for addressing the research question and whether the investigators have used an appropriate methodology in reaching the study conclusions. The checklist was written in the form of 27 questions to guide decision makers as they consider the database, the study methodology, and the study conclusions. For many of the questions, key references are provided as a resource for those who want to further examine a particular issue.

Why Would a Retrospective Database be Used for a Health-Related Research Study?

An important strength of most retrospective databases is that they allow researchers to examine medical care utilization as it occurs in routine clinical care. They often provide large study populations and longer observation periods, allowing for examination of specific subpopulations. In addition, retrospective databases provide a rela-

Address correspondence to: Brenda Motheral, PhD (Chair), Vice President, Express Scripts, 13900 Riverport Drive, Maryland Heights, MO 63043. E-mail: bmotheral@express-scripts.com

tively inexpensive and expedient approach for answering the time-sensitive questions posed by decision makers. Two recent studies have suggested that adequately controlled observational studies produce results similar to randomized controlled trials [1,2].

How Should the Checklist Be Used?

This checklist was developed primarily for the commonly used medical claims or encounter-based databases but could potentially be used to assess retrospective studies that employ other types of databases, such as disease registries and national survey data. The checklist is meant to serve as a supplement to already available checklists for economic evaluations [3,4]. Only those issues that are unique to database studies or are particularly problematic in database research were included in the checklist. Not every question will be applicable to every study. As is true with any scale or other measure of study quality or validity, the checklist cannot discern whether something was done in a particular study versus whether it was reported.

In summary, this checklist should serve as a general guide, recognizing that follow-up with study authors may be warranted when no or unsatisfactory answers to checklist questions are extant.

Data Sources

Relevance: Have the Data Attributes Been Described in Sufficient Detail for Decision Makers to Determine Whether There Was a Good Rationale for Using the Data Source, the Data Source's Overall Generalizability, and How the Findings Can Be Interpreted in the Context of Their Own Organization?

Any given database represents a particular situation in terms of study population, medical benefits covered, and how services are organized. To appropriately interpret a study, key attributes should be described, including the sociodemographic and health-care profile of the population and limitations on available services, such as those imposed by socialized medicine, plan characteristics, and benefit design (e.g., physician reimbursement approach, cost sharing for office visits, drug exclusions, mental health carve-outs). For example, in an economic evaluation that compares two drugs, it would be important to know the formulary status of the drugs as well as any other pharmacy benefit characteristics that could affect the use of the drugs, such as step therapy, compliance programs, and drug utilization review programs.

Reliability and Validity: Have the Reliability and Validity of the Data Been Described, Including Any Data Quality Checks and Data Cleaning Procedures?

With any research data set, quality assurance checks are necessary to determine the reliability and validity of the data, keeping in mind that reliability and validity are not static attributes of a database but can vary dramatically depending on the questions asked and analyses performed. Quality checks are particularly important with administrative databases from health-care payers and providers because the data were originally collected for purposes other than research, most often for claims processing and payment. Services may not be captured in the claims database because the particular service is not covered by the plan sponsor or because the service is "carved-out" and not captured in the data set (e.g., mental health). Data fields that are not required for reimbursement may be particularly unreliable. Similarly, data from providers who are paid on a capitated basis often have limited utility because providers are infrequently required to report detailed utilization information. Changes in reporting/coding over time can result in unreliable data as well. The frequency with which particular codes are used can change over time as well, often in response to changes in health plan reimbursement policies.

For all these reasons, investigators should describe the quality assurance checks performed and any steps taken to normalize the data or otherwise eliminate data suspected to be unreliable or invalid, particularly when there is the potential to bias results to favor one study group over another (e.g., outliers). The authors should describe any relevant changes in reporting/coding that may have occurred over time and how such variation affects the study findings. Data quality should be addressed even when the data have been preprocessed (e.g., grouped into episodes) prior to use by the researcher. Examples of important quality checks include missing and out-of-range values, consistency of data (e.g., patient age), and claim duplicates. Other examples of approaches that can be used to address the quality of a database are to compare data figures to established norms (e.g., rates of asthma diagnosis compared to prevalence figures) and to cite previous literature in which the database's reliability and validity have been examined [5].

Linkages: Have the Necessary Linkages among Data Sources and/or Different Care Sites Been Carried Out Appropriately, Taking into Account Differences in Coding and Reporting Across Sources?

Various types of linkages can be necessary for working with claims data. In some cases, a researcher may want to combine data from several health plans for analysis and should describe how inconsistencies in coding and reporting across health plans were addressed. For example, as new procedures or services are introduced, health plans often create their own codes so that those delivering the services can be paid. These "temporary" codes can differ across data sources, leading to variations in how the same events are reported. As to reporting, one simple scenario occurs when groups of providers, who have different relationships to the health plan, report office visits at different rates due to reimbursement arrangements. In other cases, data from one health plan may not be integrated, requiring the researcher to link all relevant health services (e.g., outpatient, inpatient, mental health, pharmaceutical, laboratory, eligibility). A particular challenge in this situation is ensuring that the each individual's records are accurately matched across data sources. This linkage process should be described, with note made of any problems that could affect data validity or study findings.

Eligibility: Have the Authors Described the Type of Data Used to Determine Member Eligibility?

In studies designed to examine outcomes over a particular time period at the patient level, it is important to determine whether patients were eligible to receive benefits during the time period. There are various types of data and approaches that might be used to determine eligibility, each with potential advantages and disadvantages, making it important that the author describe how eligibility was determined. A not uncommon but flawed approach to eligibility that is seen in the literature is the use of a prescription claim during a particular month as evidence of eligibility during that month. Because a significant percentage of members will not have a prescription claim in any given month for which they are eligible, this is an inappropriate approach to eligibility determination.

Methods

Research Design

Data analysis plan: was a data analysis plan, including study hypotheses, developed a priori? Because of the retrospective nature and relatively easy access of claims data, the opportunities for unsystematic data exploration are significant. Accordingly, it is particularly important that evidence of a well-developed a priori data analysis plan be noted for hypothesis-testing studies. For research funded by government or other nonprofit agencies, the proposal has typically undergone a rigorous peer-review process prior to funding. When other or no funding is extant, it may be unclear whether the analysis plan was developed a priori unless the authors explicitly make this statement. Hypothesis-generating studies allow for more latitude on this issue.

Design selection: has the investigator provided a rationale for the particular research design? Many designs are available to the investigator, each with particular strengths and weaknesses depending on setting, research question, and data. The investigator should provide a clear rationale for the selection of the design given the salient strengths and weaknesses of the design.

Research design limitations: did the author identify and address potential limitations of that design? Have the investigators described the potential biases, such as selection, history, maturation, and regression to the mean, and how these potential biases will be addressed?

Treatment effect: for studies that are trying to make inferences about the effects of an intervention, does the study include a comparison group and have the authors described the process for identifying the comparison group and the characteristics of the comparison group as they relate to the intervention group? If the investigation attempts to make inferences about a particular intervention, a design in which there is no comparison or control group is rarely adequate. Without a comparison group (persons not exposed to an intervention), there often exist too many potential biases that could otherwise account for an observed "treatment" effect. The comparison group should be as similar to the intervention group as possible, absent the exposure to the intervention. A rationale should be provided for selecting individual observations to the comparison group. The validity of a reported treatment effect depends on the design selected, how similar the comparison is to those exposed to the treatment, and the statistical analyses used (see "Statistics" section) [5–7].

Study Population and Variable Definitions

Sample selection: have the inclusion and exclusion criteria and the steps used to derive the final sample from the initial population been described? The inclusion/exclusion criteria are the minimum rules that are applied to each potential subject's data in an effort to define a population for study. Has a description been provided of the subject number for the total population, of the sample, and after application of each inclusion and exclusion criterion? In other words, is it clear who and how many were excluded and why? Was there a rationale and discussion of the impact of study inclusion and exclusion criteria on study findings, because the inclusion/exclusion criteria can bias the selection of the population and distort the applicability of the study findings?

Eligibility: are subjects eligible for the time period over which measurement is occurring? Databases only capture information for those patients who are "eligible" for coverage by the payer whose data are being analyzed. Hence, it is important that subjects actually be eligible to receive benefits with the payer during the time period over which they are being observed. In some cases, it may be essential that only subjects who are continuously eligible for the entire study period be included (e.g., analysis of medication continuation rates). In other cases, subjects may only be eligible for selected months during the study period, but any outcome measures (e.g., prescription claims) must be adjusted for the months of eligibility.

Censoring: were inclusion/exclusion or eligibility criteria used to address censoring and was the impact on study findings discussed? Censoring or the time limits placed at the beginning or end of the study period may potentially distort the selection and generalizability of a cohort. The investigator may choose to include only subjects who have some fixed duration of eligibility (e.g., 1 year) after the intervention. This method of right censoring (follow-up time) may bias the study if duration of eligibility is related to other factors, such as general health. For example, in government entitlement programs where eligibility is determined monthly, limiting the study population to only those with continuous eligibility would tend to include the sickest patients, as they would most likely remain in conditions that make them eligible for coverage. Alternatively, an investigator may want to identify newly treated patients and require that subjects be eligible for some period prior to use of the medication of interest. This type of left censoring should also be acknowledged and implications for study findings should be discussed.

Operational definitions: are case (subjects) and end point (outcomes) criteria explicitly defined using diagnosis, drug markers, procedure codes, and/or other criteria? Operational definitions are required to identify cases and end points, often using ICD-9-CM codes, medication use, procedure codes, etc., to indicate the presence or absence of a disease or treatment. The operational definition(s) for all variables should be provided [8].

Definition validity: have the authors provided a rationale and/or supporting literature for the definitions and criteria used and were sensitivity analyses performed for definitions or criteria that are controversial, uncertain, or novel? Investigators attempting to identify group(s) of persons with a particular disorder (Alzheimer's disease) that has some diagnostic or coding uncertainty should provide a rationale and, when possible, cite evidence that a particular set of coding (ICD-9-CM, CPT-4, Drug Intervention) criteria are valid. Ideally, this evidence would take the form of validation against a primary source but more often will involve the citation of previous research. When there is controversial evidence or uncertainty about such definitions, the investigator should perform a sensitivity analysis using alternative definitions to examine the impact of these different ways of defining events. Sensitivity analysis tests different values or combinations of factors that define a critical measure in an effort to determine how those differences in definition affect the results and interpretation. The investigator may choose to perform sensitivity analyses in a hierarchical fashion or "caseness" where the analysis is conducted using different definitions or levels of certainty (e.g., definite, probable, and possible cases).

For economic evaluations, a particularly challenging issue is the identification of disease-related costs in a claims database. For example, when studying depression, does one include only services with a depression ICD-9-CM, those with a depression-related code (e.g., anxiety), or all services regardless of the accompanying diagnosis code? As mentioned above, sensitivity analyses of varying operational definitions are important in these situations.

Timing of outcome: is there a clear temporal (sequential) relationship between the exposure

and outcome? Does the author account for proximity of key interventions to the actual event (outcome) of interest and duration of the intervention? For example, if attributing emergency room visits to use of a medication, did the emergency room visit occur during or within a clinically reasonable time period after use of the medication? One option is to create a variable for the duration (or cumulative) in time or dose and another variable that reflects the time elapsed between the most proximal intervention and the outcome itself.

Event capture: are the data, as collected, able to identify the intervention and outcomes if they actually occurred? Some procedures may not be routinely captured in claims data (e.g., office stool guiac tests) or may not be reimbursed by the payer (e.g., over-the-counter medications, out-of-network use) and thereby not captured. Such a lack of data can be an issue not only for case and end point identification but also for appropriate costing of resources in economic evaluations.

Disease history: is there a link between the natural history of the disease being studied and the time period for analysis? The researcher must address the pros and cons of the database in the context of what is known about the natural history of the disease. For example, a large proportion of the utilization for hepatitis occurs beyond the initial year of diagnosis, typically up to 10 to 20 years after diagnosis. Failing to account for this long follow-up or simply assuming a cross-section of patients adequately represents the natural history of the disease is inappropriate.

Resource valuation: for studies that examine costs, have the authors defined and measured an exhaustive list of resources affected by the intervention given the perspective of the study and have resource prices been adjusted to yield a consistent valuation that reflects the opportunity cost of the resource? Reviewers should ensure that the resource costs included in the analysis match the responsibilities of the decision maker whose perspective is taken in the research, because generally, patients, insurers, and society are responsible for paying a different set of costs associated with the intervention. For example, if the study is from the perspective of the insurer, the resource list should only include those resources that will be paid for by the insurer, which would exclude noncovered services (e.g., over-the-counter medications).

With respect to measurement, the resource use described in these data is limited by the extent of the insurance coverage. The clearest example of this is the lack of prescription utilization data for Medicare beneficiaries, because Medicare does not cover most outpatient prescriptions. This problem also occurs under insurance products where portions of benefits are carved out (e.g., mental health carve-outs) and in capitated arrangements with providers who are not required to submit detailed claims to the insurer.

Likewise, the resource should be valued in a manner that is consistent with the perspective. Typically, claims data provide a number of cost figures, including submitted charge, eligible charge, amount paid, and member copay. The perspective of the study will determine which cost figure to use. For example, if the study is from the perspective of the insurer, the valuation should reflect the amount paid by the plan sponsor, not the submitted or eligible charge.

With this being the case, the resource price information available within retrospective databases might provide an imperfect measure of the actual resource price because reported plan costs may not reflect additional discounts, rebates, or other negotiated arrangements. These additional price considerations can be particularly important for economic evaluations of drug therapies, where rebates can represent a significant portion of the drug cost. In addition, prices will vary over time with inflation and across geographic areas with differences in the cost of living. In most cases, prices can be adjusted to a reference year and place using relevant price indexes [9].

Statistics

Control variables: if the goal of the study is to examine treatment effects, what methods have been used to control for other variables that may affect the outcome of interest? One of the greatest dangers in retrospective database studies is incorrectly attributing an effect to a treatment that is actually due, at least partly, to some other variable. Failure to account for the effects of all variables that have an important influence on the outcome of interest can lead to biased estimates of treatment effects, which are referred to as a confounding bias. For example, a study might find that the use of COX-2 inhibitors is associated with subsequent higher rates of gastrointestinal (GI) events compared to NSAID users. If physicians are more likely to prescribe COX-2 inhibitors to patients with a history of GI disease and the study does not

control for the history of GI disease, then confounding basis is present. Two common approaches for addressing confounding bias in the analysis include: 1) the stratification of the sample by different levels of the confounding variables with comparison of the treatments within potential confounders (e.g., age, sex); and 2) the use of multivariate statistical techniques that allow for the estimation of the treatment effect while controlling for one or more confounders simultaneously. Each of these approaches has strengths and weaknesses.

Often investigations will attempt to control for comorbidities and or disease severity using risk adjustment techniques (e.g., Chronic Disease Score, Charlson Index). The risk adjustment model should be suitable for the population/disease that is being investigated, and a rationale for the selection of the risk adjustment model should be described [10–15].

In addition, in certain situations researchers can use methods (e.g., instrumental variable techniques) that group patients in a manner that is related to treatment choice but theoretically unrelated to unmeasured confounders. These approaches can be thought of as ex post randomizing methods, and consistent estimates of treatment effects are obtained by comparing treatment and outcome rates across groups [16].

Statistical model: have the authors explained the rationale for the model/statistical method used? Statistical methods are based on a variety of underlying assumptions. Often these stem from the distributional characteristics of the data being analyzed. As a result, in any given retrospective analysis, some statistical methods will be more appropriate than others. Authors should explain the reasons why they chose the statistical methods that were used in the analysis. In particular, the approach to addressing skewed data, a common issue in claims database research, should be described (e.g., log-transformation, two-part models).

For studies that combine data from several databases, the authors should describe what analyses have been performed to account for hierarchical or clustered data. For example, with data pooled across plans, patients will be grouped within health plans, and the health plan may have a significant impact on the outcome being measured. Outcomes may be attributed to a particular patient-level intervention, when in fact the outcome may be due to differences in health plans, such as formularies and copay amounts. Methods such as hierarchical linear modeling may be appropriate when using pooled data, and authors should discuss this issue when describing the selection of statistical methods.

Influential cases: have the authors examined the sensitivity of the results to influential cases? The results of retrospective database studies, particularly analyses of economic outcomes, can be very sensitive to influential cases. For example, an individual who is depressed and attempts to commit suicide might have extremely high medical costs that could dramatically change conclusions about the costs of treating a patient with a particular antidepressant therapy. Such "outliers" can be particularly problematic if the sample is small. There are a variety of tests to measure the sensitivity of findings to influential cases but, basically, the idea is to see how much the results change when these cases are removed from the analysis. Logarithmic transformations, commonly used to reduce the skewness in economic outcome variables, can create serious problems in making inferences about the size of statistical differences in the original (unlogged) dollar units.

Alternatively, analyses can be conducted on measures of underlying service utilization (e.g., numbers of office visits) rather than the dollar values themselves; service utilization measures tend to be less skewed than their economic counterparts. Using this approach, any identified differences in service utilization can be subsequently valued using an appropriate fee schedule. A caveat with using service utilization directly is that statistical analyses, such as regression modeling, may require the use of more sophisticated methodologies (e.g., count models) than those commonly used in expenditure analyses [17,18].

Relevant variables: have the authors identified all variables hypothesized to influence the outcome of interest and included all available variables in their model? Retrospective databases are often convenience data sets that were constructed for a purpose completely unrelated to the research study being conducted (e.g., the processing of medical claims). Although they can be extremely rich, such databases often lack information on some of the variables that would be expected to influence the outcome measure of interest. For example, the medication that a patient receives is likely to be partly a function of their clinical characteristics (primary diagnosis, medical comorbidities) and partly a function of physician prescribing patterns.

Often retrospective data sets contain information on one of these components but not the other. This is a problem because omitted variables can lead to biased estimates for the variables that are included in the model. In the special case where the omitted variables are correlated with both the treatment selection and the outcome of interest, the problem is known as selection bias. Several statistical procedures have been developed that attempt to test for, and reduce, the bias introduced by unobservable variables [19–23].

Testing statistical assumptions: do the authors investigate the validity of the statistical assumptions underlying their analysis? Any statistical analysis is based on assumptions. For example, regression analyses may include testing for omitted variables, simultaneity of outcomes and covariates, correlation among explanatory variables, and a variety of others. To have confidence in the author's findings, model specification tests should be discussed [24,25].

Multiple tests: if analyses of multiple groups are carried out, are the statistical tests adjusted to reflect this? The more statistical tests one conducts, the greater the likelihood that a "statistically significant" result will emerge purely by chance. Statistical methods have been developed that adjust for the number of tests being conducted. These methods reduce the likelihood that a researcher will identify a statistically significant finding that is due solely to chance [26–28].

Model prediction: if the authors utilize multivariate statistical techniques in their analysis, do they discuss how well the model predicts what it is intended to predict? Numerous approaches, such as goodness of fit or split samples, can be used to assess a model's predictive ability. For example, in ordinary least squares regression models, the adjusted R^2 (which measures the proportion of the variance in the dependent variable explained by the model) is a useful measure. Nonlinear models have less intuitive goodness-of-fit measures.

Models based on microlevel data (e.g., patient episodes) can be "good fits" even if the proportion of the variance in the outcome variable that they explain is 10% or less. In fact, models based on microlevel data that explain more than 50% of the variation in the dependent variable should be viewed with suspicion [29].

Discussion/Conclusions

Theoretical Basis: Have the Authors Provided a Theory for the Findings and Have They Ruled out Other Plausible Alternative Explanations for the Findings?

The examination of causal relationships is a particular challenge with retrospective database studies because subjects are not randomized to treatments. Accordingly, the burden is on the author to rule out plausible alternative explanations for the findings when examining relationships between two variables. This requires a consideration of the type of study, its design and analysis, and the nature of the results.

Practical versus Statistical Significance: Have the Statistical Findings Been Interpreted in Terms of Their Clinical or Economic Relevance?

In retrospective database studies, the sample sizes are often extremely large, which can render potentially unmeaningful differences to be statistically significantly different. In some studies that have relatively small sample sizes, the large variance in cost data can render meaningful differences statistically insignificant. Accordingly, it is imperative that both statistical and clinical or economic relevance be discussed.

Generalizability: Have the Authors Discussed the Populations and Settings to Which the Results Can Be Generalized?

While retrospective database studies often have greater generalizability than randomized controlled trials, this generalizability cannot be assumed. The authors should be explicit as to which populations and settings the findings can be generalized. In addition, the impact of changes in the health-care environment during and since the conduct of the study on generalizability should be discussed. For example, economic evaluations are sometimes conducted shortly after a product is launched, when it has not reached full market penetration. In those cases, patients studied may be systematically more or less severe than the ultimate population of users of that medication, which can impact effectiveness and cost outcomes.

We recognize the efforts of Fredrik Berggren, James Chan, Sueellen Curkendall, Bill Edell, Shelah Leader, Marianne McCollum, Newell McElwee, and John Walt, reference group members who provided comments on earlier drafts.

Travel funding for the task force meeting was provided by the ISPOR.

References

1 Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observation studies, and the hierarchy of research designs. N Engl J Med 2000;342:1887–92.
2 Benson K, Hartz AJ. A comparison of observation studies and randomized, controlled trials. N Engl J Med 2000;342:1878–86.
3 Clemens K, Townsend R, Luscombe F, et al. Methodological and conduct principles for pharmacoeconomic research. Pharmacoeconomics 1995;8:169–74.
4 Weinstein M, Siegel JE, Gold MR, et al. Recommendations of the panel on cost-effectiveness in health and medicine. J Am Med Assoc 1996; 276:1253–8.
5 McGlynn EA, Damberg CL, Kerr EA, Brook RH. Health information systems design issues and analytic applications. Santa Monica: Rand Health, 1998.
6 Campbell S, Stanley J. Experimental and quasi-experimental design for research. Chicago: Rand McNally, 1963.
7 Cook T, Campbell S. Quasi-experimentation. Chicago: Rand McNally, 1979.
8 Motheral BR, Fairman KA. The use of claims databases for outcomes research: rationale, challenges, and strategies. Clin Ther 1997;19: 346–66.
9 Lave JR, Pashos CL, Anderson GF, et al. Costing medical care: using administrative data. Med Care 1994;32(Suppl):JS77–89.
10 Ash AS, Ellis RP, Pope GC, et al. Using diagnoses to describe populations and predict costs. Health Care Fin Rev 2000;21:7–28.
11 Clark DO, Von Korff M, Saunders K, et al. A chronic disease score with empirically derived weights. Med Care 1995;33:783–95.
12 Deyo AR, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol 1992; 45:613–9.
13 Gilmer T, Kronick R, Fishman P, Ganiats TG. The Medicaid Rx model: pharmacy-based risk adjustment for public programs. Med Car 2001;39: 1188–02.
14 Lezzoni L, Ash AS, Daley J, et al. Risk Adjustment for Measuring Healthcare Outcomes (2nd ed). Chicago: Health Administration Press, 1997.
15 Kronick R, Gilmer T, Dreyfus T, Lee L. Improving health-based payment for Medicaid beneficiaries. CDPS Health Care Fin Rev 2000;21:29–64.
16 Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. J Am Stat Assoc 1996;91:444–54.
17 Mullahy J. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. J Health Econ 1998;17:247–81.
18 Cameron A, Trivedi P. Regression Analysis of Count Data. New York: Cambridge University Press, 1998.
19 Crown W, Obenchain R, Englehart L, et al. Application of sample selection models to outcomes research: the case of evaluating effects of antidepressant therapy on resource utilization. Stat Med 1998;17:1943–58.
20 D'Agostino R. Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med 1998;17:2265–81.
21 Heckman JJ. The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. Ann Econ Social Measur 1976;5:475–92.
22 Jones A. Health econometrics. In: Culyer AJ, Newhouse JP, eds., Handbook of Health Economics. Amsterdam: Elsevier, 2000.
23 Terza J. Estimating endogenous treatment effects in retrospective data analysis. Value Health 1999;2:429–34.
24 Belsley D, Kuh E, Welsh R. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: Wiley, 1980.
25 Godfrey L. Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches. New York: Cambridge University Press, 1991.
26 Tukey JW. The Problem of Multiple Comparisons [unpublished notes]. Princeton: Princeton University, 1953.
27 Scheffe H. A method for judging all contrasts in the analysis of variance. Biometrika 1953;40: 87–104.
28 Miller RG Jr. Simultaneous Statistical Inference. New York: Springer-Verlag, 1981.
29 Greene W. Econometric Analysis (4th ed.). Englewood Cliffs: Prentice Hall, 1999.