Retrospective Analyses of Large Medical Databases: What Do They Tell Us?

RICHARD A. WARD* and MICHAEL E. BRIER*†

*Division of Nephrology, Department of Medicine, School of Medicine, University of Louisville and †Veterans Affairs Medical Center, Louisville, Kentucky.

The clinical study of human disease is complicated by interdependent variables, and powerful analytical tools are necessary to establish causal relationships. Prospective studies can be randomized and blinded to the investigators. These techniques protect prospective trials, somewhat, from the problems of bias in the study design and confounding by codependent variables. However, prospective clinical trials are difficult to perform, may require extended duration for adequate observations of human diseases, and are expensive to organize and perform. Retrospective studies may take less time and are less expensive than prospective studies because the data have already been measured. However, retrospective studies are susceptible to bias in data selection and analysis. Furthermore, confounding variables may go unrecognized because of inadequate knowledge of how they interrelate with the outcome of interest. Because of these limitations, retrospective data analysis may show associations among variables, but rarely establishes causal relationships.

Dialysis has achieved widespread clinical acceptance because its efficacy is undisputed and the outcome without therapy is obvious. Patients with end-stage renal disease (ESRD) die of the complications of uremia, unless they are dialyzed or receive a renal transplant. Because the consequence of not dialyzing a patient with ESRD is so clear, dialysis was never subjected to the rigors of a prospective, randomized clinical trial. A lack of rigorous evaluation continued to characterize the development of renal replacement therapy for many years, although there were some exceptions, such as the National Cooperative Dialysis Study, which evaluated the effects of treatment time and urea removal on outcomes in hemodialysis (1). More often, retrospective, or a few prospective, clinical studies, usually in small numbers of patients, were all that were used to support changes in therapy.

More recently, the recognition that dialysis patients in the United States have much higher mortality rates than patients in Europe and Japan (2) has led to a vigorous debate about the adequacy of dialysis practices in the United States. As part of this debate, there has been renewed interest in comparing clinical outcomes between different types of renal replacement therapy in order to determine if one therapy is superior to another. Because of the perceived urgency of the question, and because of the large number of patients and long time frames required to obtain sufficient statistical power in prospective, randomized clinical studies, many of these comparisons have been performed by retrospective observational analyses of large databases.

The establishment of the United States Renal Data System (USRDS) has greatly facilitated retrospective analyses of outcomes in ESRD patients. The USRDS database contains patient-specific and center-specific data on essentially all ESRD patients treated in the United States, including demographic and medical information and ESRD treatment history (3). The USRDS presents summary statistics of these data annually and makes data files available to researchers who wish to test specific hypotheses. One example of an analysis of the USRDS database is the comparison between outcomes in peritoneal dialysis and hemodialysis reported by Vonesh and Moran in this issue of the Journal (4).

Retrospective analyses of large databases, including the USRDS database, have sometimes produced contradictory, and even controversial, results. One example of these conflicting results is the purported association between dialyzer reuse and survival in hemodialysis patients. Based on an analysis of the USRDS database, Held *et al.* (5) and Feldman *et al.* (6) concluded that certain reuse practices were associated with an increased risk of mortality. However, a recent analysis of the same data, supplemented by data from other sources, calls into question this conclusion, and suggests that the apparent relationship between dialyzer reuse and mortality is confounded by other factors, such as dialysis therapy and anemia correction (7). The current study by Vonesh and Moran (4) is another example of conflicting results arising from different analyses of large ESRD databases. Although Vonesh and Moran show no difference in outcome between peritoneal dialysis and hemodialysis, other studies have shown hemodialysis to be associated with better (8) or worse (9) outcomes than peritoneal dialysis. The failure to obtain consistent findings when different investigators analyze the same, or a similar, database raises the question of the validity of such retrospective, observational studies. Because the results of these studies may influence national treatment practices, it is important to understand their limitations and clinical usefulness.

There are few, if any, databases for other disease states comparable to the USRDS database for ESRD. As a result, one problem in assessing the results of an analysis of the USRDS database is a lack of experience and criteria with which to judge the methods of analysis. In their study, Vonesh and Moran used the USRDS database to compare outcomes between peritoneal dialysis and hemodialysis (4). Summary mortality data for five overlapping 3-year cohorts were extracted from the USRDS database, and death rates were compared between the two therapies for each cohort period. This approach, which uses summary data from a previous analysis performed by the USRDS, has certain similarities to a meta-analysis, and the rules governing meta-analyses offer one set of criteria with which to judge the validity of studies of the USRDS database (Kasiske [10] recently described the use of meta-analyses as a tool in nephrology research).

A meta-analysis may be used when experience at multiple centers has shown that the sample size needed to answer the question is not available in any one study or as a means to summarize the results of multiple studies. For example, a reviewer of the literature may attempt to combine studies in order to test the hypothesis that there is no difference between two treatments. For a meta-analysis to be valid there should be no publication bias in the combined data. All relevant data that have been generated, or a representative sample of the data, should be included in the meta-analysis and there should be no relationship between the findings in individual studies and whether or not a study was published. Furthermore, the results of the individual studies should be homogeneous and have a common dependent variable, or end point. Finally, the quality of the data being combined should be similar among studies. Because of the similarities between meta-analyses and at least some of the retrospective analyses of the USRDS database, it is reasonable to ask how well the latter satisfy the criteria for a valid meta-analysis.

Publication bias could occur if, for instance, only those studies with positive results were published and those that showed no difference between treatments were not published. In the analysis of Vonesh and Moran (4), publication bias is unlikely since all the data reported to the USRDS were used in the analysis. However, a type of publication bias may occur in database studies if particular groups are omitted from the analysis. For example, in the study of outcomes in patients treated with reused dialyzers, Held and colleagues chose to omit all patients treated in hospital-based facilities from their primary analysis (5). A subsequent analysis of the excluded data yielded a result opposite to that obtained in the primary analysis. This type of "publications bias" may limit the ability to generalize findings and, in the case of the reuse studies, may have contributed to the confusion that continues to plague this issue (6,7,11). There are other forms of publication bias, such as a language, database, or multiple publication bias (12); however, these forms of bias are unlikely to occur in analyses of the USRDS database.

Studies that are to be combined for analysis need a common end point. In the analysis of Vonesh and Moran (4), the end point is death. However, in other studies of this type, the end point may not be as clear. If the end point is hospitalization, the data may be confounded by the heterogeneity of reasons for hospitalization. For example, hospitalizations related to blood access may not be equivalent to hospitalizations related to cardiovascular morbidity, and the degree of equivalence may vary from study to study.

Finally, the quality of the combined data needs to be similar. An expert review of the data by someone familiar with the area is the usual means of ensuring comparable data quality. In the case of the USRDS database, the data are collected in the same manner for all patients. One important source of these data is the regional ESRD networks. The networks collect patient-specific data from individual dialysis centers using instruments such as the Health Care Financing Administration (HCFA) forms 2728 and 2746. The networks then compile these data and forward them to the USRDS. Some centers may be more precise than others in collecting and reporting data to their network. Unlike normal controlled clinical trials, there is very little, if any, validation of the data submitted to the USRDS database. Recently, Longnecker and colleagues examined the sensitivity and specificity of the comorbidity data reported on HCFA form 2728 (13). They found that the form 2728 was specific, but that the sensitivity was highly variable (ranging from 17% for arrhythmia to 78% for hypertension). Furthermore, the sensitivities for two comorbidities, myocardial infarction and insulin dependence, differed significantly between patients treated with peritoneal dialysis and patients treated by hemodialysis. Thus, in spite of uniform methods of data collection, the quality of the data in the USRDS database may be variable. This potential variability also extends to special studies conducted by the USRDS, such as the Dialysis Morbidity and Mortality Study (DMMS), which also rely on data submitted by the staff of individual dialysis facilities. Not only was there no validation of these data, but the staff of each dialysis facility was permitted to make inferences about the data in the absence of explicit statements in the medical record.

Based on the conditions for a valid meta-analysis, the greatest concern about analyses of the USRDS database is the quality of the data. Coupled with this general concern are some issues particular to the USRDS database. The nature of the data acquired by the USRDS has changed over time. Initially, data were retrieved only on patients whose care was eligible for reimbursement by Medicare. Patients under 65 years of age treated in-center did not become eligible for Medicare reimbursement until 90 days after the initiation of dialysis. Thus, any in-center patient under 65 years of age who died in the first 90 days of therapy was excluded from the database. In contrast, a patient under 65 years of age who chose treatment by any form of home dialysis was eligible for Medicare coverage from the day of initiation of dialysis. Thus, any home dialysis patient under 65 years of age who died in the first 90 days of dialysis was included in the database. In reporting mortality data, the USRDS considers only patients who survive through the 90th day of treatment. A revised version of HCFA form 2728 (the Medical Evidence Form) was introduced in 1995. With the introduction of the revised form, dialysis facilities were required to complete and submit the form on all patients initiat-

ing dialysis, without regard to their eligibility for Medicare. The revised form also included, for the first time, information on patient comorbidities. The failure to include all patients in the database prior to 1995, together with the problem of differences in entitlement for in-center and home dialysis patients under 65 years of age, has the potential to introduce bias into any analysis. As a result, care must be used when comparing data from USRDS reports published in different years.

In addition to changes in the nature of the data collected in the USRDS database, there have been changes in the way summary data are reported. For example, in the analysis of Vonesh and Moran (4), cohorts from 1987–1989 and 1988–1990 contained only prevalent patients; however, both prevalent and incident patients were included in the 1989–1991, 1990–1992, and 1991–1993 cohorts.

In evaluating the study of Vonesh and Moran (4), and any other studies involving the USRDS database, we must ask whether these changes in data collection and reporting violate our data assumptions and, if they do, whether the size of the change is significant. If we decide that the data are adversely affected by these changes, we then need to decide whether it would be more appropriate to analyze a subset of the data. Furthermore, we need to consider whether it is valid to compare the results of studies in which different types of data were used. The study by Bloembergen et al. used point prevalent data (9); the Canadian study used incident data (8); and Vonesh and Moran used point and period prevalent data (4). Use of all incident data may provide a different answer to these questions, as reported in a recent abstract by Collins et al. (14). This latter analysis used data from the same source as Vonesh and Moran for incident patients in the years 1994 to 1997. From these data, Collins and colleagues concluded that peritoneal dialysis was comparable to, or better than, hemodialysis except in older female diabetic patients. The exclusion of incident patients from the point prevalent studies appears to be a source of bias.

Reconciling the differences between these studies is further complicated because practice patterns have changed for both peritoneal dialysis and hemodialysis, and these changes have occurred within different time frames. For example, the debate over differences in mortality between the United States and Europe and Japan focused attention on the amount of dialysis delivered to hemodialysis patients (15,16). As a result, the delivered dose of dialysis for hemodialysis steadily increased during the early 1990s (15). In contrast, less attention has been given to the delivered dose of dialysis for peritoneal dialysis, in part because of the limited data with which to judge what constituted an adequate dose for peritoneal dialysis (17). Thus, in the time frame covered by the study of Vonesh and Moran (4), the delivered dose of hemodialysis was likely to have increased, whereas that for peritoneal dialysis is likely to have remained unchanged. This change may have biased their results in favor of hemodialysis, especially since Held and colleagues have shown that a 0.1 increase in Kt/V for hemodialysis results in a 7% reduction in mortality (16). A second major change in practice patterns over the years covered by many analyses of the USRDS database is the use of recombi-

nant human erythropoietin to correct anemia. Anemia may be predictive of an increased risk of mortality in some patients (18). Thus, analyses of morbidity may be confounded by changes in the degree of anemia, particularly since there are differences in erythropoietin and iron administration between hemodialysis and CAPD patients. Other changes in practice patterns, such as changes in provider characteristics, also have occurred, with an unpredictable impact on outcomes.

Finally, in evaluating the growing literature on outcomes in ESRD patients based on the analysis of large databases, we must consider the clinical significance of the findings. In a large database with as many observations as the USRDS database, a very small difference between two groups may be statistically significant, but the magnitude of this difference may be clinically irrelevant. Vonesh and Moran discuss this issue with the data in their Figure 2 (4). When one looks at the 1-year survival for both peritoneal dialysis and hemodialysis in the cohort years 1990–1992 and 1991–1993, we see the same survival numbers for peritoneal dialysis and hemodialysis of 0.83 and 0.82, respectively. This difference is not statistically different in the 1990–1992 cohort but is significant in the 1991–1993 cohort.

Given all these potential weaknesses, is there any utility to the type of analysis performed by Vonesh and Moran and other investigators? We need to remember that retrospective observational analyses can only show us if there is an association between variables; they cannot show causality. Nevertheless, these studies help narrow the scope of the subsequent prospective, randomized studies needed to rigorously test a hypothesis. For example, the reports of Vonesh and Moran (4) and Collins et al. (14) show an association between treatment with peritoneal dialysis, rather than hemodialysis, and decreased survival for female diabetic patients above the age of 50 to 55 years. Based on this observation, a randomized prospective study to compare outcomes between hemodialysis and peritoneal dialysis could be limited to this subset of ESRD patients.

Even using the results of the retrospective observational study to limit the scope of the subsequent randomized, prospective study, the logistics of such a study may make it difficult to perform. The current HEMO study, which is examining the impact of dose of dialysis (Kt/V) and membrane flux on outcomes in hemodialysis patients, is a good example of the resources needed to address such questions (19). Moreover, a randomized, prospective study comparing outcomes between hemodialysis and peritoneal dialysis would be complicated by the question of what is an adequate level of dialysis for both hemodialysis and peritoneal dialysis.

If it is not possible to conduct a randomized prospective trial, how much weight should we place on the results of the retrospective analyses? At present there is not a satisfactory answer to this question. We do not yet have enough experience to know which types of retrospective analysis have the greatest validity. Until there is consensus on this issue, we should be cautious. If several retrospective analyses using different methodologies and different data sets show consistent associations, and these associations agree with the results of small-scale prospective randomized studies, we may have some confidence

in the results. However, if different studies provide different results, we should be very cautious and consider which factors, not included in the analysis, may have a greater bearing on outcomes. We agree with Vonesh and Moran when they state that: "In the absence of a randomized trial comparing PD with HD, larger studies with more careful attention to case-mix adjustment, including adjustments for disease severity, dose of dialysis, nutritional status, and quality of life, are needed to better ascertain if differences in outcome between patients treated with PD *versus* HD truly exist."

References

1. Lowrie EG, Laird NM, Henry RR: Protocol for the National Cooperative Dialysis Study. *Kidney Int* 23[Suppl 13]: S11–S18, 1983

2. Held PJ, Brunner F, Odaka M, Garcia JR, Port FK, Gaylin DS: Five-year survival for end-stage renal disease patients in the United States, Europe and Japan, 1982 to 1987. *Am J Kidney Dis* 15: 451–457, 1990

3. U.S. Renal Data System: *USRDS 1998 Annual Data Report*, Bethesda, MD, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, April 1998

4. Vonesh EF, Moran J: Mortality in ESRD: A reassessment of differences between patients treated with hemodialysis and peritoneal dialysis. *J Am Soc Nephrol* 10: X-X, 1999

5. Held PJ, Wolfe RA, Gaylin DS, Port FK, Levin NW, Turenne MN: Analysis of the association of dialyzer reuse practices and patient outcomes. *Am J Kidney Dis* 23: 692–708, 1994

6. Feldman HI, Kinosian M, Bilker WB, Simmons C, Holmes JH, Pauly MV, Escarce JJ: Effect of dialyzer reuse on survival of patients treated with hemodialysis. *JAMA* 276: 620–625, 1996

7. Collins AJ, Ma JZ, Constantini EG, Everson SE: Dialysis unit and patient characteristics associated with reuse practices and mortality: 1989–1993. *J Am Soc Nephrol* 9: 2108–2117, 1998

8. Fenton SSA, Schaubel DE, Desmeules M, Morrison HI, Mao Y, Copleston P, Jeffery JR, Kjellstrand CM: Hemodialysis versus peritoneal dialysis: A comparison of adjusted mortality rates. *Am J Kidney Dis* 30: 334–342, 1997

9. Bloembergen WE, Port FK, Mauger EA, Wolfe RA: A comparison of mortality between patients treated with hemodialysis and peritoneal dialysis. *J Am Soc Nephrol* 6: 177–183, 1995

10. Kasiske BL: Meta-analysis as a clinical tool in nephrology. *Kidney Int* 53: 819–825, 1998

11. Kimmel PL, Mishkin GJ: Dialyzer reuse and the treatment of patients with end-stage renal disease by hemodialysis. *J Am Soc Nephrol* 9: 2153–2156, 1998

12. Egger M, Smith GD: Meta-analysis: Bias in location and selection of studies. *Br Med J* 316: 61–66, 1998

13. Longnecker JC, Klag MJ, Coresh J, Levey AS, Martin AA, Fink NE, Powe NR: Validation of comorbid conditions on the ESRD medical evidence report by medical record review: The choices for healthy outcomes in caring for ESRD (CHOICE) study [Abstract]. *J Am Soc Nephrol* 9: 218A, 1998

14. Collins A, Ma J, Xia H, Ebben J: CAPD/CCPD in incident patients is equal to or better than hemodialysis, except in females ages ≥55 years [Abstract]. *J Am Soc Nephrol* 9: 204A, 1998

15. Owen WF, Lew NL, Liu Y, Lowrie EG, Lazarus JM: The urea reduction ratio and serum albumin concentration as predictors of mortality in patients undergoing hemodialysis. *N Engl J Med* 329: 1001–1006, 1993

16. Held PJ, Port FK, Wolfe RA, Stannard DC, Carroll CE, Daugirdas JT, Bloembergen WE, Greer JW, Hakim RM: The dose of hemodialysis and patient mortality. *Kidney Int* 50: 550–556, 1996

17. *NKF-DOQI Clinical Practice Guidelines for Peritoneal Dialysis Adequacy*, New York, National Kidney Foundation, 1997, pp 134–137

18. Madore F, Lowrie EG, Brugnara C, Lew NL, Lazarus JM, Bridges K, Owen WF: Anemia in hemodialysis patients: Variables affecting this outcome predictor. *J Am Soc Nephrol* 8: 1921–1929, 1997

19. Eknoyan G, Levey AS, Beck GJ, Agodoa LY, Daugirdas JT, Kusek JW, Levin NW, Schulman G: The hemodialysis (HEMO) study: Rationale for selection of interventions. *Semin Dial* 9: 24–33, 1996