Outcomes Research in Nephrology

James E. Rohrer

The field of clinical research in nephrology should be and is being broadened beyond clinical trials by incorporating epidemiologic designs and subjective patients' assessments of their own health and the degree to which they are satisfied with services they receive. The advantages of these changes include increased relevance to the concerns of patients and increased relevance to the real world of clinical practice.

THE FIELD OF OUTCOMES research lies in the overlap between clinical research and health services research. Health services research is the study of the structures, processes, and outcomes of health care. The term *clinical research* usually seems to equate with drug trials in the minds of many medical faculty, but clinical epidemiology certainly should be included. Health services research includes medical economics, health psychology, medical sociology, and even health care management. For the purposes of this discussion, I will focus on patient health outcomes, rather than all of the myriad consequences of health care. That subset of health services research that addresses quality of care and clinical effectiveness is most relevant to clinical research. For our purposes, outcomes research is both clinical epidemiology and health services research.

Several methodologic issues must be considered when studying the outcomes of care. These include the validity and appropriateness of the outcome measures, the design of the study, and how to adjust for clinical and other differences among the patients. Published reports from the nephrology literature are used to show several of the observations made.

MEASUREMENT OF OUTCOMES

Mortality is the most common outcome measure seen in the nephrology literature. Assuredly, mortality is important. However, a single-minded focus on mortality could have negative consequences for the quality of care from a patient's perspective. Therefore, outcome studies ideally would include multiple types of outcome measures, each reflecting a different type of goal.

Consider the following vignette: Mrs. Jones is 89 years of age, a widow, in very poor overall health, and living in a nursing home. She is somewhat confused, but usually able to articulate her wishes. Last year, she was placed on a 3 times per week dialysis program because of the certainty of death without it. She was very unhappy with dialysis and complained frequently; however, her children insisted that she continue. Bedridden and weak, she occasionally would resist caregivers. For the last few months, she has been plucking out her dialysis shunts when not restrained.

Mrs. Jones is still alive, so her nephrologist might congratulate himself on a successful case. But of course Mrs. Jones' satisfaction with care and quality of life would be very low. Although some readers might regard this vignette as unbelievable, there is no doubt that nephrologists can be found who believe and say that if they can keep a person alive then the case is a success. If the patient is not happy with life quality, then he or she may be advised that his or her attitude could be improved.

Clearly, health-related quality of life and satisfaction with services are legitimate and important outcome measures, even if they rarely are addressed. The medical outcome study led to the development of the short form-36 (SF-36), which was intended for use over time to chart a patient's progress. A large quantity of information on reliability and validity has been produced about the SF-36.[1] However, 36 questions make for a questionnaire of impractical length in many situations. The developers of the SF-36 also produced the SF-12, which also is valid, but recommended for measuring differences between populations rather

Department of Health Services Research and Management, Texas Tech University Health Sciences Center, School of Medicine, Lubbock, TX.

Address reprint requests to James E. Rohrer, PhD, Professor and Chair, Department of Health Services Research and Management, Texas Tech University Health Sciences Center, School of Medicine, 3601 4th St, MS 8161, Lubbock, TX 79430. Email: jrohrer@ama.ttuhsc.edu

0270-9295/03/23003-0002$30.00/0
doi:10.1016/S0270-9295(03)00060-3

than changes in the health of an individual subject. Both the SF-36 and the SF-12 are copyrighted.

The Centers for Disease Control and Prevention (CDC) uses 4 questions in a scale that measures health-related quality of life. These questions included self-rated health, number of days out of the past 30 when physical health was not good, number of days when mental health was not good, and number of days when normal activities could not be performed. Validity information has been provided on this scale as well.[2] The CDC questions are not copyrighted.

Obviously, the CDC index measures health status, and by implication the CDC has concluded that health status and health-related quality of life are overlapping constructs. If we are just measuring health status, then why not use single items to measure physical and mental health? That in fact has been performed. *Health Services Research*, which is a flagship journal in health services research, recently published as its lead article a study of self-rated health and self-rated depression, each measured with a single question.[3] Another recent study used self-rated health and a single mental health item to study community health in Vancouver.[4] Self-rated health has been shown to be a good predictor of subsequent mortality.[5] Psychometricians may not be comfortable with this kind of reductionism, but epidemiologists regard exchanging 36 questions for 2 as an unqualified success, provided that those 2 items will do the job that is expected of them.

A similar story can be told about measurement of satisfaction with care. Instruments can become quite long if all aspects of care are assessed. For example, Rubin et al[6] found 18 broad aspects of dialysis care, ranging from satisfaction with the nurse to satisfaction with the nephrologist. However, this level of detail is not always necessary. Overall satisfaction with care is a legitimate outcome and one that patients can judge accurately with a single item.

At the opposite extreme from self-ratings of life quality and satisfaction are clinical measures that may be termed *biomarkers*. For example, Donadio et al[7] studied the effects of dietary fish oil on the progression of renal disease. Outcome measures included serum creatinine and 24-hour urine protein measurements. Harris et al,[8] in a study of the effects of multidisciplinary case management, used serum creatinine level and creatinine clearance as outcome measures. Although biomarkers are important measures, they are less important than how well the patient can go about his or her daily activities.

The best approach to measurement of outcomes might follow decision rules such as the following: (1) measure self-rated health and mental health and describe them as health status measures; (2) collect appropriate biomarkers; and (3) use a disease-specific quality-of-life measure that is based on the symptoms of the disease or the disability it causes.

The foregoing discussion has stressed the importance of adding subjective measures to outcome studies so as to broaden the definition of effectiveness to include the patient's perspective. However, it is important to recognize that self-ratings are not precise variables; large numbers of cases are needed to fairly test the hypothesis that different treatment approaches have different outcomes. A study of 40 subjects, not uncommon in the clinical literature, will not be sufficient. Depending on the measure used, a few hundred cases may be required to have adequate statistical power.

DESIGN OF STUDIES

The assumption of this article is that observational epidemiology actually is the most practical approach to outcomes research for most clinicians. The randomized experiment is of course the most reputable approach to research and has the strongest internal validity. Consequently, many physicians and scientists disregard observational studies. Some have been heard to say observational studies are not research because they are not valid—a patently false conclusion but one that is logical given the strong bias in favor of randomized studies found in the medical literature. However, randomized studies have flaws not found in other designs. For example, they often analyze subjects who may not be typical of those found in ordinary clinical practice. Furthermore, they analyze treatment delivered under very rigorous and therefore atypical circumstances. Poor adherence to protocols by patients and clinicians is prevented in clinical trials, but is commonplace in the real world. The results of randomized experiments may be misleading. Often what works in the experiment does not work as expected in normal clinical practice. And randomization does not always accom-

plish the purpose of eliminating important differences between groups.

A 1999 study that used a retrospective design revealed some of the limitations of randomized study designs.[9] The investigators were interested in the prognosis of patients with type II diabetes who were entering dialysis, which they described as abysmal in the title of their article. Data on 84 consecutive patients was abstracted, who were divided into 3 groups: acute, chronic, and acutely aggravated chronic renal failure. Most patients began dialysis under emergency conditions. Mean follow-up time was 211 days, at which time 32% of these subjects were found to have died, mostly from cardiovascular diseases. The investigators concluded that the "factors aggravating renal failure were mainly iatrogenic, and therefore largely avoidable."[9] These iatrogenic problems included poor treatment for the cardiovascular diseases at the back end, and late referral at the front end.

The issue of late referral is important. The nephrologist may assume that late referral is beyond his control or responsibility, but if better coordination with primary care can make a difference, then perhaps the issue of responsibility should be reconsidered. Patients should be under the care of a nephrologist for a full year before dialysis begins, yet this appears to happen less than 50% of the time.[10] From an outcomes research perspective, this means that most studies have used biased samples of subjects. For the most part we do not know what outcomes will be experienced by patients if we start nephrology care a year earlier then is now the case. Although we might suspect that outcomes will improve, a researcher would want to test that hypothesis. Nevertheless, the evidence to date seems to support the conclusion that late referral reflects poor predialysis care, suggesting that improved coordination of providers and better education of patients are in order.[11]

One implication of this line of reasoning is that a dialysis start can be regarded as a negative outcome.[12] As indeed it is, from the patient's point of view.

Obviously, a tightly controlled study with rigid protocols would not reveal iatrogenic contributors to poor outcomes, such as late referral. Even so, the preceding argument should not be taken as a recommendation that randomized studies should never be performed. Such studies have their place, and it is an important one. A clinical trial is appropriate to answer questions about whether one treatment is better than another when both are performed as they should be.

CASE MIX ADJUSTMENT

When analyzing the outcomes of care, the investigator must always consider the possibility that unmeasured differences among patients may be the reason for differences in outcomes. In fact, the search for patient characteristics that predetermine outcomes is an important ongoing activity that more clinical outcomes researchers should consider. Most will continue to focus on the relative effectiveness of treatments, but results of treatment studies could be debunked easily owing to omission of an important risk factor. Indeed, the rhetoric of epidemiology with its focus on independent risk factors is important here. For example, Yu et al[13] reported that predialysis glycemic control is an independent predictor of clinical outcome in type II diabetic patients who are on continuous peritoneal dialysis. They classified 60 patients into good or poor glycemic control. Patients were drawn from the dialysis unit of a single university hospital over an 8-year period. The results of this study were interesting: even though no significant differences in predialysis morbidity could be shown, survival was better for patients whose predialysis glycemic control was better. This would appear to be useful information to have in hand when studying survival rates among dialysis patients.

Observational studies must be very concerned about the possibility that unmeasured patient differences will confound the results. However, randomized studies also should measure possible confounders and test for the possibility that they explain the findings. For example, Mehta et al[14] compared continuous and intermittent dialysis for acute renal failure. Over 160 patients were randomized into the 2 types of treatment. The mortality rate was found to be higher for continuous therapy. However, significant differences were found between the groups despite randomization. Gender, hepatic failure, Acute Physiology and Chronic Health Evaluation (APACHE) II and II scores, and the number of failed organ systems were related independently to mortality and also occurred differently between the 2 treatment groups. After adjusting for these differences, the

odds of death were revealed to be not significantly different between the 2 groups.

Adjusting for patient differences is critical, yet the appropriate method for doing so is subject to debate. Fiaccadori et al[15] reviewed and evaluated 3 severity-of-illness scoring systems for use when studying the outcomes of acute renal failure: APACHE II, version II of the Simplified Acute Physiology Score (SAPS II), and version II of the Mortality Probability Model at 24 hours (MPM_{24} II). They concluded, after studying 425 patients admitted for acute renal failure, that all 3 scoring systems predict mortality with some accuracy, but none is accurate for individual patients. For research purposes we might say that the ability to detect differences between groups at baseline is the critical purpose for severity measures. All 3 of these scoring systems will perform this function.

Alternatively, the investigator might consider comparing groups at baseline on the variables used to compute severity. The APACHE uses physiologic variables, premorbid major organ dysfunction (scaled from 1–5), and age. The SAPS uses physiologic variables, age, type of admission (scheduled surgical, unscheduled surgical, or medical) and the presence of acquired immune deficiency syndrome, metastatic cancer, and hematologic malignancy. The MPM uses physiologic variables, chronic diagnosis, acute diagnosis, type of admission, age, use of cardiopulmonary resuscitation, and use of mechanical ventilation as risk variables. If all of these variables were tested separately to verify similarity of groups, then only those that were found to be significantly different could be used to adjust outcome variables. After all, if the requisite data are to be collected anyway for computation of a severity score, then they might as well be used individually.

CONCLUSIONS

Measurement of self-rated physical and mental health status is an important adjunct to outcome studies in nephrology. Such measures admittedly are subjective. On the one hand, capturing the patient's perspective requires a willingness to use subjective measures. Satisfaction with care also is subjective. Patients' evaluations of care do not accurately reflect the technical skill and knowledge used by their physicians. However, patient evaluations are the ultimate measure of whether physicians are meeting the expectations of the people whom they serve.

In addition to collection of subjective data, outcome researchers in nephrology should be encouraged to consider use of observational designs, including cohort studies, case-control studies, and, occasionally, cross-sectional surveys. Observational studies require that patient characteristics that may affect outcomes be measured and the outcomes must be adjusted for those patient differences. Analysis of subjective data and use of observational designs will require sample sizes that are somewhat larger than some investigators would prefer.

However, nephrology will not be able to improve outcomes for its patients until it begins to routinely measure the spectrum of results experienced by its patients. Furthermore, results as they occur in the real world of clinical practice are of more immediate relevance to patients than outcomes measured under artificial conditions. The future direction of outcomes research in nephrology will incorporate these twin priorities into its armamentarium: subjective data from patients and observational designs.

REFERENCES

1. McHorney CA, Ware JE Jr, Lu JF, et al: The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 32:40-66, 1994

2. Centers for Disease Control and Prevention: Measuring Healthy Days. Atlanta, GA, Centers for Disease Control, 2000

3. Shi L, Starfield B, Politzer R, et al: Primary care, self-rated health, and reductions in social disparities in health. Health Serv Res 37:529-550, 2002

4. Dunn JR: Housing and inequalities in health: A study of socioeconomic dimensions of housing and self reported health from a survey of Vancouver residents. J Epidemiol Community Health 56:671-681, 2002

5. Burstrom B, Fredlund P: Self rated health: Is it as good a predictor of subsequent mortality among adults in lower as well as in higher social classes? J Epidemiol Community Health 55:836-840, 2001

6. Rubin HR, Jenckes M, Fink NE, et al: Patient's view of dialysis care: Development of a taxonomy and rating of importance of different aspects of care. CHOICE study. Choices for Healthy Outcomes in Caring for ESRD. Am J Kidney Dis 30:793-801, 1997

7. Donadio JV Jr, Grande JP, Bergstralh EJ, et al: The long-term outcome of patients with IgA nephropathy treated with fish oil in a controlled trial. Mayo Nephrology Collaborative Group. J Am Soc Nephrol 10:1772-1777, 1999

8. Harris LE, Luft FC, Rudy DW, et al: Effects of multidis-

ciplinary case management in patients with chronic renal insufficiency. Am J Med 105:464-471, 1998
9. Chantrel F, Enache I, Bouiller M, et al: Abysmal prognosis of patients with type 2 diabetes entering dialysis. Nephrol Dial Transplant 14:129-136, 1999
10. Levin A: Consequences of late referral on patient outcomes. Nephrol Dial Transplant 15:8-13, 2000 (suppl 3)
11. Arora P, Obrador GT, Ruthazer R, et al: Prevalence, predictors, and consequences of late nephrology referral at a tertiary care center. J Am Soc Nephrol 10:1281-1286, 1999
12. Levin A, Lewis M, Mortiboy P, et al: Multidisciplinary predialysis programs: Quantification and limitations of their impact on patient outcomes in two Canadian settings. Am J Kidney Dis 29:533-540, 1997
13. Yu CC, Wu MS, Wu CH, et al: Predialysis glycemic control is an independent predictor of clinical outcome in type II diabetics on continuous ambulatory peritoneal dialysis. Perit Dial Int 17:262-268, 1997
14. Mehta RL, McDonald B, Gabbai FB, et al: A randomized clinical trial of continuous versus intermittent dialysis for acute renal failure. Kidney Int 60:1154-1163, 2001
15. Fiaccadori E, Maggiore U, Lombardi M, et al: Predicting patient outcome from acute renal failure comparing three general severity of illness scoring systems. Kidney Int 58:283-292, 2000