

Special Article

Treatment Research at the Crossroads: The Scientific Interface of Clinical Trials and Effectiveness Research

Kenneth B. Wells, M.D., M.P.H.

Objective: Policy and clinical management decisions depend on data on the health and cost impacts of psychiatric treatments under usual care, i.e., effectiveness. Clinical trials, however, provide information on treatment efficacy under best-practice conditions. An understanding of the design, analysis, and conventions of both efficacy and effectiveness studies can lead to research that better informs clinical and societal questions. **Method:** This paper contrasts the strengths and limitations of clinical trials and effectiveness studies for addressing policy and clinical decisions. These research approaches are assessed in terms of outcomes, treatments, service delivery context, implementation conventions, and validity. **Results:** Clinical trials and effectiveness research share problems of internal and external validity despite more attention to internal validity in clinical trials (e.g., randomization, blinding, standardized protocols) and to external validity in effectiveness studies (e.g., community-based treatments, representative samples). **Conclusions:** To develop research at the interface of clinical trials and effectiveness studies, research goals must be redefined, and methods, such as cost-utility and econometric analyses, must be shared and developed. Development of hybrid designs that combine features of efficacy and effectiveness research will require separation of conventions such as frequency of follow-up, intensity of measurement, and sample size from the central scientific issues of aims and validity.

(*Am J Psychiatry* 1999; 156:5–10)

The gold standard for obtaining information on how psychiatric treatments affect health has been the randomized controlled clinical trial. Clinical trials evaluate efficacy, i.e., whether a given treatment per se improves outcomes relative to a control or comparison condition. Achieving this goal often requires testing treatments under ideal or best-practice conditions. This represents an important step toward determining the desirability of implementing a treatment in practice. Yet clinical trials, owing to features of design and implementation, have important limitations for in-

forming clinical practice and policy decisions about treatments. In particular, patients and providers are especially interested in effects of treatments as delivered in the community, outside of rigorous clinical trials, i.e., effectiveness. Further, clinical trials often focus on relatively discrete choices, such as treatment modality or intensity, rather than practical management decisions, such as whether to hospitalize a patient or use physical restraints. Findings from clinical trials can inform policy debates concerning the design of mental health benefits by clarifying the potential therapeutic value of treatment, but policy debates increasingly require information that is directly generalizable to community patient samples (e.g., users in an insurance plan) and that relates to outcomes of societal consequence, such as long-run morbidity and costs.

To develop such information, we need a broad research agenda that includes efficacy studies, effectiveness studies, and hybrid studies that use features of

Received Oct. 10, 1997; revisions received March 5 and June 23, 1998; accepted Aug. 25, 1998. From the Neuropsychiatric Institute and Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles. Address reprint requests to Dr. Wells, UCLA Neuropsychiatric Institute, Suite 300, 10920 Wilshire Blvd., Los Angeles, CA 90024-6505; kwells@ucla.edu (e-mail).

Supported in part by NIMH grants MH-54623 and MH-01170.

both. This article explores some of the key differences in perspectives and methods of efficacy and effectiveness research and separates underlying scientific issues and research conventions. The article was written by a mental health services researcher, and efficacy researchers may be more skeptical of the effectiveness approach.

My central thesis is that efficacy and effectiveness studies rely on different prevailing design strategies and analysis approaches and that they often have competing implementation conventions. As a result, there is no quick fix that transforms one kind of research into the other. An understanding of both approaches and uncoupling of conventions from scientific issues can lead to studies that better inform clinical and societal questions. My view is also that the emergence of an effectiveness perspective reflects an underlying paradigm shift toward greater concern with societal impacts of treatment and toward a corresponding reexamination of what is considered relevant scientific evidence on the value of treatments.

DEFINING THE INTERFACE

Efficacy studies examine whether treatments improve outcomes under controlled conditions that optimize isolation of the treatment effect through design features, such as a control or placebo condition, randomization, standardized treatment protocols, homogeneity of samples, and blinding of subjects, providers, and evaluators (1, 2). Case-control and historical comparisons are also used in the development and testing of therapies. Clinical trials often entail substantial deviations from usual practice conditions, by eliminating treatment preferences, providing free care, using specialized providers and settings, maintaining high treatment compliance, and excluding patients with major comorbid conditions.

Effectiveness studies evaluate effects of treatments on health outcomes under conditions approximating usual care (3). There is no agreement over which features of usual care define an effectiveness perspective, however. I suggest that an effectiveness study should evaluate a treatment that is feasible for community application, include community treatment settings, and rely on representative patients or providers in these settings. Cost-effectiveness studies evaluate the marginal difference in cost for a marginal difference in outcome for one treatment relative to an alternative. Cost-effectiveness studies are particularly important for informing policy decisions (4) but are much less common than efficacy or effectiveness studies.

Effectiveness studies are more heterogeneous in design than clinical trials. Some are controlled experiments (5, 6), but such experiments often eliminate or modify design features of clinical trials that protect internal validity, such as blinding or treatment standardization. Effectiveness studies commonly use quasi-experimental designs (7). Some treatment effectiveness

analyses occur in the context of a larger health services study of differences in financing or organization of health care delivery. The Medical Outcomes Study, for example, matched providers by specialty, patients by medical "tracer" conditions, and service systems by location to compare patient outcomes in prepaid and fee-for-service care (8). Observational effectiveness studies examine natural variations in exposure to treatments and rely on statistical techniques to adjust for baseline patient differences in comparing outcomes. Sturm and Wells (9) used the Medical Outcomes Study data as observational data, stratifying by initial sickness and imputing treatment effects through a decision analysis.

OUTCOMES

While efficacy and effectiveness studies can include similar outcomes, clinical trials in psychiatry are usually designed to evaluate short-term clinical outcomes while effectiveness studies are more often designed to evaluate long-term clinical and morbidity outcomes. Similarly, the clinical detail tends to be much greater in efficacy studies, and the cost and morbidity detail greater in effectiveness studies. Some effectiveness analyses, particularly those that are part of larger service delivery studies, rely on proxy health outcomes, such as readmission rates (10, 11), limiting their utility for informing clinical practice. Disease-specific clinical outcomes, such as course of disorder, are common in clinical trials and can provide useful information for practice and some policy debates. But clinical and policy decisions regarding resources for different disease conditions or for mental health versus social programs, require data on outcomes, such as morbidity, that apply across conditions. While it is becoming common to include morbidity measures, such as the Medical Outcomes Study 36-item Short-Form Health Survey (SF-36), in clinical trials, this practice may not help inform policy or management debates except in the context of meta-analyses. Morbidity outcomes have high variance, and studies of them require much larger samples than are typical of psychiatric clinical trials, i.e., fewer than 100 subjects.

The recommended outcome for cost-effectiveness studies is health utility, or preference for health states (4). Utilities integrate diverse outcomes into a single score, permitting comparisons across diverse treatments and disorders and offering a singular "bottom line" to policy debates. Utilities have rarely been applied in psychiatric treatment studies (12). Assessment of utilities is controversial and technically challenging, but this method would allow more effective public debates on the value of psychiatric treatments (4, 13–15). Cost-effectiveness studies also require adequate assessments of costs. Direct costs are the costs of treatment and changes in health care costs. Societal costs are changes in productivity and use of human resources related to treatment. Assessment of costs is complex be-

cause there are many components, some of which, such as use of general medical providers for mental health care, are difficult to assess reliably (16). Further, costs are highly variable and the sample sizes required for treatment comparisons are often at least 200 subjects per cell. Meta-analysis may allow cost-effectiveness analyses across multiple clinical trials (17). An example is the cost-utility meta-analysis by Kamlet et al. (12) for maintenance treatment of recurrent major depression.

Policy makers are particularly interested in programs that affect societal productivity over many years, but few psychiatric studies of either type have had this broad scope.

TREATMENTS

Treatment studies have limited usefulness if the treatments are not feasible in community practice or their relationship to usual care is unknown. For example, while the literature supports the efficacy of structured forms of psychotherapy for major depression (13, 18, 19), the relationship of these therapies to community practice is unclear, limiting the usefulness of efficacy findings to debates about psychotherapy coverage. Potential solutions to the problem include identifying community therapies that are equivalent to efficacious treatments, evaluating the effectiveness of community therapies, or standardizing community therapy to approximate efficacious approaches. Studies of psychotropic medications have a qualitatively similar problem because the adherence rates achieved in practice are lower than in clinical trials. The solution in this case requires knowing how noncompliance affects effectiveness and monitoring compliance rates in practice.

In many effectiveness studies, treatments or programs are more of a "black box" than in efficacy studies, which typically rely on manual-based treatment protocols. Effectiveness studies may have greater heterogeneity or individual variation in treatment approaches, which may be hard to describe or assess except crudely. For example, "counseling for depression" in the Medical Outcomes Study was defined as any discussion of depression for at least 3 minutes during a medical visit (8). Some effectiveness studies of community treatments, such as assertive community treatment (20, 21), use manual-based protocols or compare usual care to efficacious therapies. The Patient Outcomes Research Team for depression, funded by the Agency for Health Care Policy and Research, uses a protocol to encourage but not require primary care providers to follow guideline-based recommendations in treating depressed patients with medications or psychotherapy (22). A major challenge in effectiveness research is to standardize interventions while preserving usual-care conditions.

SERVICE DELIVERY CONTEXT

Because a main goal of effectiveness research is to determine how treatments work when applied in practice, these studies require a description of the clinical setting or service delivery context. However, there are no standards for such a description, and many effectiveness studies lack even basic information. A minimal standard could include a description of the training level or specialty mix of providers, provider-staff ratios, availability of resources (programs, ancillary staff) to support mental health care, mix of financing strategies (e.g., percent capitated care), and presence or absence of clinical management structures, such as utilization review and quality assurance programs. Reports of clinical trials in psychiatry often have no description of the service delivery context, perhaps because the subjects are not necessarily recruited from a given health care system and treatment is provided under study-specific conditions. However, investigators in both types of studies should more systematically document the service delivery context and attempt to identify contextual factors likely to affect treatment or outcomes.

For some clinical management and policy purposes, it is important to understand how features (i.e., organization and financing) of the health care delivery affect patient outcomes through differences in rates of treatment, i.e., a structural model. This purpose exceeds the scope of treatment effectiveness per se and instead represents the intersection of effectiveness and health services or policy research. Examples are the Health Insurance Experiment (23) and the Minnesota Medicaid Capitation Trial (24), both of which provided evidence for worse mental health outcomes in fee-for-service cost-sharing plans (relative to free-care plans) or in capitation plans (relative to fee-for-service plans) among the poor with severe psychopathology. In the Prospective Payment System Quality of Care Study, outcomes of inpatient management of psychotropic medication for depressed elderly patients were examined (11) as one component of an evaluation of the impact on quality of care of Medicare's prospective payment system, which is based on diagnosis-related groups. Pursuing such an agenda requires linking multiple levels of data and developing targeted opportunities for very large interdisciplinary studies.

IMPLEMENTATION CONVENTIONS

There are major differences in implementation conventions between effectiveness and efficacy studies. Relative to clinical trials in psychiatry, effectiveness studies, particularly when embedded in a policy study, often have larger samples. The typical psychiatric clinical trial has 20–100 subjects, and the NIMH Treatment of Depression Collaborative Research Program (25), a very large clinical study, had 250 subjects. The effectiveness studies of Katon et al. (5, 6) and Schul-

berg et al. (26) were only slightly smaller (153–217 subjects), while the Medical Outcomes Study (8) and the depression Patient Outcomes Research Team (22) are 3–5 times as large. As noted earlier, effectiveness studies may have less clinical detail than do efficacy studies. For example, the NIMH collaborative study of depression treatment (25) had follow-up sessions that lasted several hours and assessed multiple comorbid psychiatric conditions, while the Medical Outcomes Study (8) had brief self-report assessments of about an hour and assessed only selected comorbid disorders. Some effectiveness studies, particularly those conducted as part of larger health services studies, have longer durations of follow-up, e.g., 1 to 4 years (5, 6, 8) than do clinical trials, i.e., weeks or months (1). Clinical trials typically have more frequent (e.g., weekly) follow-up assessments than do effectiveness studies (e.g., every few months).

Studies at the interface will increasingly require both clinical and societal perspectives (e.g., cost data), and achieving both is difficult and expensive and almost invariably leads to compromises in the depth and breadth of data collected. For example, the depression Patient Outcomes Research Team (22) has brief measures of both social costs and comorbid psychiatric conditions that may not satisfy scientists familiar primarily with either labor economic studies or clinical trials.

VALIDITY

Internal validity refers to the certainty that the study findings are true for the study population and setting. External validity refers to the generalizability of the findings to other populations and settings. The two concepts are related: generalizability is reduced in the face of poor internal validity, and high internal validity is irrelevant if the findings cannot be applied. As a general rule, efficacy studies place a higher priority on internal validity while effectiveness studies place a higher priority on external validity. However, both types of studies share threats to both types of validity.

Problems in external validity for clinical trials are related to the convenience sampling method used for recruiting subjects, the exclusion criteria, and the use of specialized samples of patients and providers and specialized treatment conditions. Problems in external validity for effectiveness studies can relate to the same factors, but effectiveness studies more often have representative sampling techniques at the patient level (8) and include at least some features of usual care conditions in the treatment protocol. However, it is relatively rare even in effectiveness studies to have representative sampling of providers and systems of care, because it is very expensive to do so and there are no national listings of all mental health providers or central directory of health care delivery systems. Developing county, state, or national data on providers and health care systems is necessary for consideration of

the generalizability of studies and for providing a sampling base for larger effectiveness studies. Further, achieving greater generalizability means obtaining the cooperation of multiple insurers and community providers and facing the many research implementation problems (27).

Clinical trials, despite randomization and blinding, can have problems with internal validity because of initial group differences despite randomization, subversion of randomization or blinding, differential refusal or dropout rates, noncompliance with treatment protocols, and contamination (crossover between treatment conditions) (28–31). The medical literature includes examples of such problems in clinical trials, and meta-analyses demonstrate that flaws in blinding are common and lead to overestimation of treatment efficacy (32–34). Effectiveness studies, especially those that are not randomized, are more likely to have initial differences in compared groups, and they share with clinical trials the problems of attrition and dropouts.

Both efficacy and effectiveness studies can have limited internal validity because of how the data are analyzed. For both types of studies, statistical techniques can be used to describe and control for bias in initial group assignment or from dropouts. Common techniques are to control for baseline sickness or rely on pre-post change scores in analyses. But there are problems in using the usual analytic techniques, such as analysis of covariance or regression techniques, to control for bias when the group differences are large or the samples are small. For example, psychiatric clinical trials with samples of under 100 subjects often have adequate power for detecting moderate to large (0.5 standard error) initial group differences in demographic and clinical variables but low power for detecting differences in either rare events (e.g., hospitalizations) or factors that have high variability (e.g., morbidity).

An important distinction is between an intent-to-treat analysis, which compares groups as initially randomized, and an as-treated analysis, which compares patients on the basis of actual treatment exposure. For example, a common practice in medication clinical trials is to exclude from analyses patients who discontinue an assigned treatment or to reassign them according to the actual treatments received. This results in observational use of the experimental data; i.e., a true intent-to-treat analysis is not possible. Observational studies usually support only as-treated analyses. In addition, many analyses in efficacy and effectiveness studies focus on nonrandomized factors, such as treatment history, as predictors of outcome (35, 36). For example, one analysis of data from a randomized clinical trial of minor tranquilizers for anxiety disorder (36) focused on the effects of prior history of tranquilizer use, not on the randomized treatment.

These features lead to two potential problems: 1) many statistical methods assume random exposure to treatment and are inappropriate for as-treated analyses, and 2) the compared groups may not be equivalent on measured and unmeasured characteristics, leading to

biased estimates of treatment effects. Measured characteristics are those assessed in the study and available for analyses. The usual approach to bias is to control analytically for initial group differences in measured characteristics. Unmeasured characteristics are unavailable for analyses but still affect treatment-outcome relationships. Reports of clinical trials in psychiatry rarely mention unmeasured factors, but they are the greater potential threat to validity. The Medical Outcomes Study affords an example of bias due to measured and unmeasured variables in analyses of treatment for depression. In unadjusted analyses, having any treatment (antidepressant medication or counseling for depression plus a series of visits) versus none was associated with significantly worse 2-year morbidity for depressed patients. In analyses adjusting for measured baseline health differences, treated and untreated patients had comparable 2-year outcomes. In analyses designed to minimize unmeasured bias by restricting the range of sickness, among patients with the most severe depression, treatment was associated with significantly better 2-year outcomes (8).

Achieving greater internal validity means implementing stronger experimental or quasi-experimental designs in both efficacy and effectiveness studies and, for larger studies, using advanced analytic techniques developed in econometrics, such as decision analysis, structural modeling, and analysis of instrumental and propensity variables (15, 30, 37–41). These techniques are relatively unfamiliar in psychiatry but have been applied and refined over the last 15 years in policy analysis. For example, instrumental variables analysis relies on identification of an extraneous factor, or “instrument,” that is randomly distributed with respect to outcomes but affects variation in exposure to treatment. The technique yields a range of likely treatment effects, given a set of assumptions about unknown factors, such as the variance of unmeasured variables. Structural modeling and propensity analyses, in contrast, do not directly address unmeasured factors. They can be used, respectively, to confirm whether data relationships are consistent with a hypothesized causal model and to match simultaneously on multiple confounding (measured) variables to efficiently estimate treatment effects. Both of these techniques are based on the assumption that all of the relevant variables have been measured. Some of these techniques have low precision and can require huge samples (thousands of subjects), as in a recent instrumental variable analysis of invasive treatments for elderly patients with acute myocardial infarction (42).

Internal and external validity can rarely both be optimized in one study, so studies at the interface of efficacy and effectiveness research will have a range of threats to validity that must be appreciated by the researcher and the reader. The conventional wisdom is that researchers should state the limitations of any study, meaning that investigators in nonexperimental studies should be particularly cautious about causal inference and all investigators should be appropriately

cautious about generalizability. But the goal of most treatment outcome studies is to inform a causal question, users of the findings may not understand the subtleties, and policies are often put in place with no relevant data; observational treatment data, if well analyzed, are much better than no data.

SUMMARY AND IMPLICATIONS

Current trends in treatment research reflect concerns with obtaining more generalizable scientific evidence on treatment effects, implementing effective treatments in community settings, and designing and evaluating treatments that enhance outcomes, such as functioning and cost-utility, that are of societal as well as individual concern. These trends are leading to applied studies with less protection of internal validity and greater focus on external validity, although true generalizability is difficult to define and achieve. These trends also imply that first-stage clinical treatment studies should focus more on the development and testing of therapies that are specifically designed to improve societal outcomes, entailing a shift in priorities for treatment development. Later-stage clinical trials will shift toward dropping exclusion criteria, expanding outcomes measures, and applying treatments in community settings. There is also a need for the goals of clinical trials to be more informed by the findings of observational studies of treatment in communities, and vice versa. Methodological advances, such as broader application of econometric methods and meta-analysis and development of national sampling frames for providers and health care systems, hold promise for advancing the field but raise further questions about feasibility and technical capacity. Both efficacy and effectiveness studies should more systematically collect both health and costs outcomes. Studies should identify factors that predict treatment compliance rather than dropping noncompliers from analyses. Both types of studies should distinguish experimental and observational uses of data and apply appropriate analytic tools and interpretive standards to each purpose. Achieving greater generalizability will require larger and more expensive studies. Engaging managed care companies, providers, and consumers in formulating this agenda will increase the relevance of treatment research to stakeholders. Integration of efficacy and effectiveness approaches into hybrid studies would be facilitated by having mental health services researchers study clinical trial protocols and design features, having efficacy researchers learn how to assess morbidity and cost outcomes and design quasi-experimental studies, and having both types of investigators become familiar with cost-effectiveness and econometric analysis methods. However, we do not yet know whether the findings of such hybrid studies will be clinically and socially useful.

The broad questions clinicians and scientists face at the interface of efficacy and effectiveness studies are the following: What scientific information about treat-

ment is in the best public interest? Can such information be used to improve individual and public health? Can we develop the research methods, training, and opportunities to achieve these goals?

REFERENCES

- Bulpitt C: *Randomized Controlled Clinical Trials*, 2nd ed. Boston, Kluwer Academic, 1996
- Fava M, Rosenbaum JF: *Research Designs and Methods in Psychiatry*. New York, Elsevier, 1992
- Gabbard GO, Lazar SG, Hornberger J, Spiegel D: The economic impact of psychotherapy: a review. *Am J Psychiatry* 1997; 154:147–155
- Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB: Recommendations of the Panel on Cost-Effectiveness in Health and Medicine. *JAMA* 1996; 276:1253–1258
- Katon W, Von Korff M, Lin E, Walker E, Simon GE, Bush T, Robinson P, Russo J: Collaborative management to achieve treatment guidelines: impact on depression in primary care. *JAMA* 1995; 273:1026–1031
- Katon W, Robinson P, Von Korff M, Lin E, Bush T, Ludman E, Simon G, Walker E: A multifaceted intervention to improve treatment of depression in primary care. *Arch Gen Psychiatry* 1996; 53:924–932
- Campbell D: *Experimental and Quasi-Experimental Designs for Research*. Boston, Houghton-Mifflin, 1963
- Wells KB, Sturm R, Sherbourne C, Meredith L: *Caring for Depression*. Cambridge, Mass, Harvard University Press, 1996
- Sturm R, Wells KB: How can care for depression become more cost-effective? *JAMA* 1995; 273:51–58
- McGlynn E, Norquist GS, Wells KB, Sullivan G, Liberman RP: Quality-of-care research in mental health: responding to the challenge. *Inquiry* 1988; 25:157–170
- Wells KB, Rogers WH, Davis LM, Kahn K, Norquist G, Keeler E, Koseoff J, Brook RH: Quality of care for hospitalized depressed elderly patients before and after implementation of the Medicare prospective payment system. *Am J Psychiatry* 1993; 150:1799–1805
- Kamlet MS, Wade M, Kupfer DJ, Frank E: Cost-utility analysis of maintenance treatment for recurrent depression: a theoretical framework and numerical illustration, in *Economics and Mental Health*. Edited by Frank RG, Manning WG Jr. Baltimore, Johns Hopkins University Press, 1992, pp 267–291
- Russell RL, Orlinsky DE: Psychotherapy research in historical perspective: implications for mental health care policy. *Arch Gen Psychiatry* 1996; 53:708–715
- Gelijms AC: *Modern Methods of Clinical Investigation*. Washington, DC, Institute of Medicine, National Academy Press, 1990
- Gelijms AC, Halm A: *The Changing Economics of Medical Technology*. Washington, DC, Institute of Medicine, National Academy Press, 1991
- Simon GE, Von Korff M, Heiligenstein JH, Revicki DA, Grothaus L, Katon W, Wangner EH: Initial antidepressant choice in primary care. *JAMA* 1996; 275:1897–1902
- Province MA, Hadley EC, Hornbrook MC, Lipsitz LA, Miller JP, Mulrow CD, Ory MG, Sattin RW, Tinetti ME, Wolf SL: The effects of exercise on falls in elderly patients: a preplanned meta-analysis of the FICSIT Trials. *JAMA* 1995; 273:1341–1347
- Depression Guidelines Panel: *Depression in Primary Care, vol II: Treatment of Major Depression: Clinical Practice Guideline 5*: AHCPR Publication 93-0551. Rockville, Md, Agency for Health Care Policy and Research, 1993
- Antonuccio DO, Danton WG: Psychotherapy versus medication for depression: challenging the conventional wisdom with data. *Professional Psychol: Res and Practice* 1995; 26:574–585
- Burns BJ, Santos AB: Assertive community treatment: an update of randomized trials. *Psychiatr Serv* 1995; 46:669–675
- McFarlane WR, Dushay RA, Stastny P, Deakins SM, Link B: A comparison of two levels of family-aided assertive community treatment. *Psychiatr Serv* 1996; 47:744–750
- Wells KB: The design of Partners in Care: evaluating the cost-effectiveness of improving care for depression in primary care. *Soc Psychiatry Psychiatr Epidemiol* (in press)
- Newhouse JP, Insurance Experiment Group: *Free for All? Lessons From the RAND Health Insurance Experiment*. Cambridge, Mass, Harvard University Press, 1993
- Lurie N, Moscovice IS, Finch M, Christianson JB, Popkin MK: Does capitation affect the health of the chronically mentally ill? results from a randomized trial. *JAMA* 1992; 267:3300–3304
- Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, Glass DR, Pilkonis PA, Leber WR, Docherty JP, Fiester SJ, Parloff MB: National Institute of Mental Health Treatment of Depression Collaborative Research Program: general effectiveness of treatments. *Arch Gen Psychiatry* 1989; 46:971–982
- Schulberg HC, Block MR, Madonia MJ, Scott CP, Rodriguez E, Imber SD, Perel J, Lave J, Houck PR, Coulehan JL: Treating major depression in primary care practice: eight-month clinical outcomes. *Arch Gen Psychiatry* 1996; 53:913–919
- Wells KB, Astrachan BM, Tischler GL, Unützer J: Issues and approaches in evaluating managed mental health care. *Milbank Q* 1995; 73:57–75
- Louis T, Shapiro S: Critical issues in the conduct and interpretation of clinical trials. *Annu Rev Public Health* 1983; 43:25–46
- Sackett D, Gent M: Controversy in counting and attributing events in clinical trials. *N Engl J Med* 1979; 301:1401–1412
- Heckman JJ: Randomization and social policy evaluation, in *Evaluating Welfare and Training Programs*. Edited by Manski CF, Garfinkel I. Cambridge, Mass, Harvard University Press, 1992, pp 201–230
- Kramer MS, Shapiro S: Scientific challenges in the application of randomized trials. *JAMA* 1984; 252:2739–2745
- Schulz KF, Chalmers I, Grimes DA, Altman DG: Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA* 1994; 272:125–128
- Schulz KF, Chalmers I, Hayes RJ, Altman DG: Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273:408–412
- Schulz KF: Subverting randomization in controlled trials. *JAMA* 1995; 274:1456–1458
- Pocock SJ: *Clinical Trials: A Practical Approach*. New York, John Wiley & Sons, 1983
- Rickels K, Case WG, Downing RW, Winokur A: Long-term diazepam therapy and clinical outcome. *JAMA* 1983; 250:767–771
- Heckman J, Singer B: *Longitudinal Analysis of Labor Market Data*. New York, Cambridge University Press, 1985
- Holland P: Statistics and causal inference. *J Am Statistical Assoc* 1986; 81:945–960
- Little RJA, Rubin DB: *Statistical Analysis With Missing Data*. New York, John Wiley & Sons, 1987
- Rosenbaum P, Rubin D: The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70:41–55
- LaLonde R: Evaluating the econometric evaluations of training programs with experimental data. *Am Economics Rev* 1986; 76:604–620
- McClellan M, McNeil B, Newhouse J: Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? analysis using instrumental variables. *JAMA* 1994; 272:859–866