Randomized Controlled Trials in Evidence-Based Mental Health Care: Getting the Right Answer to the Right Question

by Susan M. Essock, Robert E. Drake, Richard G. Frank, and Thomas G. McGuire

Abstract

The purpose of clinical research is to answer this question: Would a new treatment, when added to the existing range of treatment options available in practice, help patients? Randomized controlled trials (RCTs)—in particular, double-blind RCTs—have important methodological advantages over observational studies for addressing this question. These advantages, however, come at a price. RCTs compare treatments using a particular allocation rule for assigning patients to treatments (random assignment) that does not mimic real-world practice. "Favorable" results from an RCT indicating that a new treatment is superior to existing treatments are neither necessary nor sufficient for establishing a "yes" answer to the question posed above. Modeled on an experimental design, RCTs are expensive in time and money and must compare simple differences in treatments. Findings have a high internal validity but may not address the needs of the field, particularly where treatment is complex and rapidly evolving. Design of clinical research needs to take account of the way treatments are allocated in actual practice and include flexible designs to answer important questions most effectively.

Keywords: Randomized trial, evidence-based practice, methodology.


The Gold Standard of the Double-Blind Randomized Trial

Randomized controlled trials (RCTs) have evident methodological advantages over observational studies for assessing the impacts of treatment. Random assignment to treatment conditions means that every person participating in a study has the same chance of being assigned to a given treatment or control condition. Given a sufficient sample size, we can assume that each treatment condition will contain a comparable mix of people. If the researcher can ignore differences in the groups that arise from chance alone, any differences in outcomes between the groups can be attributed to differences in the treatments. Randomization balances the treatment and control groups on both variables that are observed (and can be controlled for statistically in naturalistic designs) and those that are unobserved (and cannot be directly controlled for in naturalistic studies). Randomized trials separate treatment effects from time-related effects. Randomization also allows us to use statistical tests that rely on independent sampling. Statistical methods such as propensity-score and instrumental-variable analyses are available to contend with nonrandom assignment, but the validity of these methods in any particular application rests on assumptions that cannot be tested directly. Furthermore, randomization is a fair way to allocate a limited resource (e.g., access to a new treatment).

Nonexperimental studies without random assignment to treatment conditions can yield misleading findings. For example, women taking hormone replacement therapy (HRT) were observed to have fewer adverse cardiac events than apparently comparable women who did not take HRT, and hence HRT was believed by many to offer some protection against heart disease. However, when women were randomly assigned to HRT or placebo, those in the HRT condition were found to be at increased risk for adverse cardiac events (Writing Group for the Women's Health Initiative Investigators 2002). Apparently, even after age and socioeconomic status were controlled for, the correlational studies comparing women who did and did not use HRT did not have comparable groups.

Observational studies and RCTs have led to different findings in psychiatry as well. When clozapine was first introduced, mirror-image studies (with baseline measures...
as each person’s control) suggested that clozapine would cause a substantial reduction in the length of inpatient stays (Mallya et al. 1992; Wilson 1992; Dennis et al. 1993; Chiles et al. 1994; Ebrahim et al. 1994; Reid et al. 1994; Wilson and Claussen 1994). In contrast, later data from randomized trials indicated that clozapine’s effects on length of stay were less than the correlational data had suggested (Essock et al. 1996a, 1996b; Rosenheck et al. 1997). The differences in findings come about because, in routine practice, treatment assignment is not random. Clinicians tend to assign treatment to those whom they believe need or would benefit from it most, introducing a correlation between treatment condition and “outcome” due to the assignment process and not due to the effectiveness of treatment. In another example, case managers spend a disproportionate fraction of their efforts on people who are not doing well, because these are the people that need their help. The apparent finding that people who get the most case management services have the poorest outcomes (even when controlling for observable covariates) does not necessarily reflect a causal relationship indicating that case management is counterproductive.

Double-blind randomized trials (DBRTs) are a subset of RCTs that keep the clinicians, patients, and raters of outcome in the dark about whether a particular person is a control subject or has been given the experimental treatment. The purpose is to avoid any conscious or unconscious bias favoring a treatment condition. The methodological advantage weakens the connection between the laboratory and the context in which the treatments will ultimately be administered. Routine practice does not blind clinicians, and as a consequence the physician can dose to the medication at hand (e.g., with respect to doses per day, and size and rate of dosage increments) rather than to a protocol. A patient would not be asked to take a pill twice a day if the medication’s recommended dosing is once a day, but this is the case in a double-blind trial comparing a once-per-day medication with a twice-per-day medication. These comments illustrate a theme of this article: DBRTs have the best chance of avoiding bias in comparisons, but the comparisons that are possible given the constraints imposed by the recruitment process, eligibility criteria, randomization, and blinding limit the external validity of these less biased comparisons. At the extreme, such trials are perfect in execution but offer only limited information regarding the impact of introducing a new treatment into practice. Clinical research on new treatments needs to answer questions without bias, but they must be the right questions—those that inform us about how introducing a new treatment will affect people’s health.

Some treatments cannot be blinded; people in randomized trials comparing a supported employment treatment program with a referral to traditional employment services know what treatment condition they are in. For trials that aim to assess an intervention’s effectiveness in close-to-actual practice environments, one option is to conduct single-blind trials where the raters, but not the patients and clinicians, are blind to an individual’s treatment assignment and other aspects of the person’s treatment. In that way, expectation effects of the individuals performing the clinical assessments are controlled but expectation effects of the patients and treatment providers are part of the impact of the treatments under study, as they are in daily clinical interactions.

Putting the Gold Standard Into Context

To understand the strengths and limitations of DBRTs, and where they fit into a program of clinical research, it is helpful to consider the process of research leading to evidence to support new treatments. The purpose of clinical evaluations of new treatments is to provide evidence on the health impacts of making a new treatment available for introduction into real-world clinical practice. Research proceeds in a sequence of steps. The first is development of evidence regarding whether a treatment is a good candidate for introduction into practice by demonstration that the treatment is “efficacious.” The research question at this stage is, Can the treatment be shown to offer clinical benefits relative to either placebo (or no treatment) or an existing treatment under “laboratory” conditions? When we refer to “laboratory conditions,” we mean control conditions that guarantee a high level of adherence to the proposed treatment protocol for populations for whom the treatment is most likely to be efficacious; in other words, conditions tending toward favorable findings for the innovative approach. The logic of beginning with research of this type is that if a new treatment fails under “ideal” conditions to improve over existing treatments, it is unlikely to be useful in clinical practice.

Assessing the potential of a new treatment theory should require a high standard of evidence that the new treatment causes an improvement in outcomes when administered under favorable conditions to a population targeted on theoretical grounds. The assessment should account for the enthusiasm of the developers of the intervention along with the hopes of patients, who often are without attractive alternatives. That such an initial test abstracts from daily practice means that treatments introduced into practice will be based on theoretical treatment technologies that have met a high initial standard of efficacy. It is here that the DBRT is ideally suited to provide the requisite evidence.

Not all efficacious treatments would necessarily improve clinicians’ ability to help patients. A second step in the research program, therefore, seeks to identify the
efficacious treatments that would improve care if they were introduced into the practice settings where the vast majority of people are treated. Efficacy research narrows the set of candidate treatments, but a key policy question remains. Would introducing treatment X for the treatment of the population for which it was designed improve care in actual practice? An “effectiveness trial” addresses part of this question. In an effectiveness trial, subjects are randomly assigned to the new treatment or an alternative, where the alternative treatment is commonly treatment-as-usual. Effectiveness trials, as distinct from efficacy research, involve fewer constraints on the administration of treatment, thereby better approximating how the new treatment would be deployed in typical practice. For example, the dosing regime may be left to the prescribing physician rather than being protocol driven—thereby, in theory, allowing finer tuning to the patient’s needs but also, in practice, allowing drift from what may be optimal prescribing. Longer trial durations help address the effectiveness of an intervention by allowing longer term risks and benefits to accrue and be measured. Patient populations studied in effectiveness trials are often less homogeneous than those studied in DBRTs or efficacy studies. They generally consist of the types of patients that the treatment was intended to serve, but, in actual practice, for medication as well as psychosocial treatments, patients deemed eligible for the new treatment may be less rigorously screened than those in an efficacy trial (e.g., screening may rely on chart diagnosis rather than on a diagnosis made on the basis of a Structured Clinical Interview for DSM-IV [First et al. 1995]). Sometimes effectiveness trials consider a wider range of outcomes than efficacy trials. These might include indicators of functioning or cost-effectiveness. The evidence generated by effectiveness trials offers insights into the impact of introducing a new treatment into typical clinical practice under situations of relatively few constraints over the way the treatment is administered, thereby yielding strong generalizability regarding the populations to whom the treatment is given.

The benefits and costs of random assignment in effectiveness trials are more complex matters than in efficacy trials. Part of the evaluation of how a treatment would be used in actual practice involves study of how clinicians and patients would sort themselves into the new treatment condition in comparison to existing treatments. Patients would certainly not be “randomly assigned” to the new treatment. The comparison of a treatment and control group in an effectiveness trial does not, therefore, squarely address the question of whether introduction of the new treatment would improve practice.

To complete the consideration of a new treatment’s impact on health outcomes, one also needs to understand through research what types of patients would receive the treatment in practice, how the treatment would be administered to these groups, and what the outcomes would be. This requires the examination of the “allocation effects” of a new treatment. Patients, providers, and care managers respond to the introduction of new treatments by changing the allocation of patients to treatments in a nonrandom fashion. A new treatment may not be used in the manner in which it is intended and is certainly unlikely to be used in a manner mimicking random assignment in an effectiveness trial. If so, the randomized effectiveness trial may not be a good guide to what to expect with the introduction of the new treatment.

We can illustrate this point with a well-known example from the mental health field. A number of randomized effectiveness trials of partial hospitalization were undertaken in the 1970s and 1980s (Horvitz-Lennon et al. 2001) showing that partial hospitalization was effective and cost-effective relative to traditional inpatient treatment (Dickey et al. 1989; Creed et al. 1990). Those studies were based on the random assignment of people at risk for hospitalization to either a partial hospital program or a psychiatric inpatient setting. Patients in the partial hospital programs achieved similar or superior outcomes at lower cost than those in the traditional inpatient treatment condition. When Medicare conducted a demonstration program and subsequently introduced partial hospital care as a covered service, the observed results were quite different (Morrison et al. 1985; Office of the Inspector General 1998). The reason for the divergence in results was that the majority of people who accessed partial hospital care were not at risk for hospitalization (i.e., they were less acutely ill than were the individuals who participated in the effectiveness trials). No offsets in use of high-intensity/high-cost treatment occurred, and no significant clinical gains were observed in the aggregate. The patients who sought out the new treatment were not at high risk for inpatient care, and thus partial hospital care represented an increase in both intensity and cost of care. The new service did not improve outcomes, but it did increase costs. Allocation effects subverted a potentially cost-effective treatment.

In stark terms, one can say that evidence from a randomized effectiveness trial is not sufficient to support the position that a new treatment would improve the effectiveness of care in practice. To make matters worse, success in a randomized effectiveness trial is also not necessary for a new treatment to improve the effectiveness of care in practice. Allocation effects can subvert the effectiveness of treatment, but they can also enhance the effectiveness of treatment over that which would be achieved by random assignment. Random assignment allocates patients between treatment and control groups in a “mindless” fashion, disregarding information potentially available to the clinician and/or patient.
about who would more likely benefit from treatment. Take a simple example: suppose a new treatment requires an extra effort from patients in terms of adherence, perhaps in the form of some activity or exercise they must undertake outside of clinical supervision. If the new treatment were given to “random” patients, it would include those patients who would and would not be able to put in the extra effort, and in comparison to the control group, the average patient with the new treatment might do worse. (This result might continue to be found even in a regression context with subgroup analyses, if the variable correlated with good outcome is difficult to measure.) If, on the other hand, clinicians were able, as they are in actual practice, to select which patients might be good candidates for the new treatment, the new treatment might be effective and cost-effective. In this example, allocation effects work in favor of the new treatment.

In sum, success in a randomized controlled effectiveness trial, because of the confounding of allocation effects, is neither necessary nor sufficient to answer the research question of whether introduction into practice would improve care. Allocation effects present a formidable challenge to assessing the potential health impacts of a new clinical intervention. Allocation effects are typically the domain of health services research, not clinical research, and the investigators skilled in one area tend not to be in the other.

DBRTs of efficacy or effectiveness are not always feasible for practical reasons. Patients may be unable to stay in the assigned treatment arm long enough for researchers to be able to make statistically valid comparisons between the subject groups as randomized. The intention to treat sometimes is just that, only an intention. For example, adolescents randomly assigned to medication or 16 weeks of psychotherapy may find one or the other objectionable and discontinue, and the reasons for and rates of discontinuation may differ across the treatment arms. When the dropout from the treatment arms is nonrandom (as is the case, for example, when individuals drop out because of medication side effects or because they find the treatment regime too burdensome), the assumptions underlying the statistical approaches commonly used to analyze data from clinical trials are violated. Statistical adjustment can partially repair the damage in such cases, but at minimum, these design problems increase the uncertainty in control-experimental comparisons.

Clinical trials for disorders such as schizophrenia are very short in comparison to the duration of the illness, and these short trials may not provide accurate estimates of the outcomes to be had from maintenance treatment. Both positive and negative outcomes may take many times the trial’s duration to accrue. Yet longer trials mean that individuals may be unwilling to follow a study protocol for the duration of the trial. Adjunctive treatments cannot be fore-stalled indefinitely, making the ability to isolate a signal from noise increasingly difficult with time. Longer trials both add to and detract from statistical power, often in ways that can be estimated only crudely.

The field of psychiatry has shifted from a model of acute care and curative interventions to one of long-term disease management and rehabilitation, with a corresponding shift in the skill sets required by clinicians (Cruz and Pincus 2002). The challenge for intervention researchers at all points on the efficacy-effectiveness continuum is to help provide the tools to meet this challenge. The National Institute of Mental Health (NIMH) has called for a bridging of science and service so that its research program will better enable people with mental illnesses to receive optimal care (National Advisory Mental Health Council 1999). Using research to inform decision making when it comes to advising state commissioners of mental health as to where to invest their mental health dollars to maximize recovery for people with serious mental disorders is a complicated matter that will not and should not turn on efficacy findings alone. Michael Hogan, the director of the Ohio Department of Mental Health and the chair of the President’s New Freedom Commission on Mental Health, has stated the challenge to researchers clearly: “Researchers and funders of research must speedily move beyond studies that cannot be generalized. . . . The evidence that increasingly counts is the evidence that treatment works for consumers in real-world settings, not for researchers in university clinics” (Hogan 2002). A friendly amendment to this statement might be that what counts are treatments that work in usual care settings and in university clinics. The intervention-research challenges associated with this shift involve identifying ways to enhance the development of lasting skills (for clinicians as well as consumers), the provision and titration of supportive services, and improving consumer’s everyday functioning and quality of life.

Asking randomized trials to incorporate these outcome measures in any meaningful way may be unrealistic because short trials cannot capture these longer-term outcomes and longer trials cannot retain enough individuals in the assigned treatment arms to allow an unbiased estimate of the effectiveness of the interventions being contrasted. For example, we may know a great deal about how to get someone a job within 18 months (a very long duration for most randomized trials) but very little about helping people with serious mental disorders develop satisfying longitudinal careers. Similarly, in studies of programs for people who are homeless and mentally ill, 30 days of stable housing is an outcome measure but it does not predict escaping from homelessness and developing long-term stable housing, which is the real goal of such treat-
ment interventions. Clearly, a research program heavy on randomized trials could be very light on information useful to line clinicians and mental health system administrators.

Effectiveness Research and Allocation Effects of a New Treatment

Finally, as a third step in a research program that begins with DBRTs or other efficacy studies, effectiveness studies can be carried out as demonstration research with the purpose of incorporating any allocation effects of the new treatment. Demonstration research is meant to mirror the “real world” as closely as possible. Can effectiveness trials be designed to better account for possible allocation effects? One possibility is to consider the rationing systems within which treatments are likely to be introduced in their original design. The field is beginning to see investigators modify existing evidence-based treatments with an eye to particular service system characteristics. Examples include recent efforts to introduce modified forms of cognitive behavior therapies for geriatric patients with anxiety disorders (Unützer et al. 1997) and to treat depressed patients in primary care settings so as to fit within the organization and management of primary care offices (Wells and Sherbourne 1999). Constructing effectiveness trials that include rationing rules limiting the scope of populations and treatment in ways that are consistent with existing formulary and utilization management arrangements is another possibility. These might include stepped care, prior authorization rules, and other similar controls. In this way, a treatment technology could be implemented in the context of specific rationing arrangements that would tend to make the trial results better approximate likely allocation effects.

Do No Harm—Where a Randomized Trial May Lead One Astray

Time and logistical constraints mean that researchers never can perform all the relevant RCTs that would be required to carry out factorial designs including all possible treatment arms crossed by all possible patient characteristics. Consider a question as simple as, When might the coprescribing of different antipsychotics (polypharmacy) be a preferable alternative to monotherapy? Given the prevalence of polypharmacy in routine practice and the paucity of research data to support this practice, this is a pressing question. The challenges to mounting studies in this area include the huge number of polypharmacy possibilities (there is no reason to assume that a given polypharmacy combination can represent all possible antipsychotic pairings) as well as the smaller, but still daunting, number of monotherapy agents to consider. Researchers in the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) study made the difficult design decision to include one first generation antipsychotic agent (perphenazine) to compare with four second generation agents. Mounting this five-arm DBRT takes a multimillion-dollar nationwide effort. Including more than one first generation antipsychotic as a treatment arm or introducing past medication history, current psychosocial treatments, current levels of social support, or past or current substance use as randomization variables would have made the study design too large and too complex to be feasible, yet we have strong research data to indicate that each of these variables may affect outcome. Randomization to treatment condition does not wash away all sins; it simply minimizes the risk of unintended differences across treatment groups at baseline. If participation in a given randomization condition is also associated with differential access to other factors that may influence outcome, then the treatments for individuals in that treatment arm may differ from other arms in significant and unmeasured ways. For example, if individuals assigned to treatment X become less assaultive, and if people who are assaultive are often excluded from vocational services, then study results with employment as an outcome may indicate that treatment X has better employment outcomes, whereas what really is happening is that individuals in treatment X, but not Y or Z, had preferential access to additional effective services. One cannot create a study design that would avoid this dilemma; the “perfect” design would be enormous and too complex to carry out. Rather, we can follow the staged data collection process described above, recognizing that, inevitably, there will be arms missing (CATIE has one first generation agent arm, rather than an arm for each possible first generation agent, and a polypharmacy arm first occurring in phase 3 of the trial design requiring that two prior treatments have failed).

In 1999, then-NIMH-director Steve Hyman presented his charge to the newly created NIMH Intervention Research Review Committee by challenging the reviewers to ask whether an application proposed the most rigorous design that would be feasible for the question being considered. This standard recognizes that the most rigorous design imaginable (e.g., a factorial design of all possible patient categories by all treatment conditions) may not be achievable, yet science will progress only if we are willing to implement studies that will still leave some questions unanswered and some variables confounded for disentangling at a later juncture if warranted based on the accumulating knowledge base. He challenged us to favor studies that would push important areas forward rather than to give the best scores to pristine designs that would offer
only narrow insights into well-trodden areas. We should not let the best be the enemy of the good, because the best may not be feasible. To criticize CATIE for having only one first generation agent as a treatment arm would be to ask for a cleaner design without regard for the difficulty of mounting a study where even one randomization arm is to a first generation antipsychotic.

In studies focusing on medication effectiveness, all possible permutations of currently available medications would be impossible to study, and the variations of available psychosocial treatments are even more complex. Even when RCTs are available, because of where and with whom they were conducted, of necessity they contain limitations related to patient characteristics, clinicians’ skill sets, service settings, and living circumstances. As noted above, data from RCTs conducted under efficacy conditions may not be generalizable to complicated, comorbid patients; more typical clinicians, often with more limited training and experience; and idiosyncratic circumstances like small rural clinics or community clinics in ethnic or gay neighborhoods. The evidence-based medicine approach assumes that there are always limitations of the evidence and that clinicians must consider a hierarchy of evidence (Guyatt and Rennie 2002). That is, a systematic review of RCTs may not always provide an unequivocal basis for decision making in a specific clinical situation. Rather, the typical clinical situation diverges in many ways from existing RCTs and may require considerable clinical judgment on the part of the clinician. As we next explain, the answer is not always to call for more and more RCTs under effectiveness conditions.

RCTs are not always necessary or appropriate. There are many clinical decisions in which the research question can be solved more easily than with an RCT and in which it is prudent to save the few potential RCTs for more complex studies. To cite one simple example, artificial hip replacement surgery for disabling arthritis was never studied with an RCT because people who were crippled began to ambulate without pain soon after surgery (cited in Guyatt and Rennie 2002). Consider a more complex example from psychiatric rehabilitation. Fifteen years ago, experts in psychiatric rehabilitation argued that an average of 6 months of counseling was required to help clients with severe mental illness make a realistic job choice. That was the prevailing theory of psychiatric rehabilitation. However, interviewing a range of clients promptly contradicted that theory. The majority of clients, even among individuals who were chronically unemployed, indicated that they had job preferences and that their preferences were realistic, at the first meeting prior to any counseling (Becker et al. 1996). Thus, it made no sense to conduct a 6-month RCT of counseling versus no counseling to study the outcome of realistic job choice. As theories and practices in the field of psychiatric rehabilitation evolved, the appropriate question for an RCT became whether pre-employment counseling prior to searching for a job is effective.

Several RCTs have addressed this question, and the consistent answer is that pre-employment counseling does not produce better employment outcomes of any kind (type of job, quality of job, amount of work, wages, satisfaction with job, etc.) (Bond 1998). Doing an RCT to study 6 months of pre-employment counseling to choose a job would have been premature and would have led to a study that was irrelevant by the time of completion.

Many complex psychosocial interventions may be difficult or impossible to study with RCTs. There are many reasons for this. Psychosocial interventions often involve combining a variety of interventions for people with complex needs who are at points in their illness where they are likely to reject treatments after brief, if any, exposure to them. The clinician’s challenge is to try to engage the patient in an array of treatments (medication, housing supports, substance abuse treatment, money management, etc.) and to tailor the approach to the needs and preferences and time of life of the patient. Therefore, the overall interventions change continually, in part to track the patient’s changing needs and preferences and in part as engagement strategies on the part of the clinician. Many individuals with schizophrenia have such complexities to their treatment needs, so inviting their participation in a treatment study means that, once they are randomly assigned to a treatment condition, they may decline the treatment being offered, they may participate only briefly, or they may drop out and decide to return or to partake of a service being offered as part of another treatment arm. All of this means that trials of long enough duration to let the treatments being considered have an effect are also long enough that there will be large dropout from, or partial adherence to, the assigned treatment conditions, resulting in nonequivalent groups and after-the-fact reduction to quasi-experimental studies. As noted previously, complex interventions may be difficult to implement in routine practice settings. Once the interventions are implemented, drift may occur easily and often.

Consider, for example, the situation with regard to studies of interventions for patients with dual diagnoses (i.e., those with co-occurring severe mental illness and substance use disorders). Most experienced clinicians would agree that the appropriate treatment for such patients involves a combination of individual, group, family, rehabilitative, self-help, and housing interventions, often supplemented by money management, laboratory testing, and expert pharmacology (e.g., Mueser et al. 1998). These different component interventions are evolving rapidly, and they are often difficult to specify and implement. Moreover, different combinations of component interventions are appropriate for different patients,
and the optimal combination for a specific patient changes over time as the patient moves through stages of recovery. Furthermore, patients move in and out of psychosocial rehabilitation treatments. The net result is that the great majority of RCTs in the field of dual diagnosis have failed or have been transformed into quasi-experimental studies (Drake et al. 1998). RCTs that have been completed are severely limited by small samples (Godley et al. 1994), highly selected compliant patients (Barrowclough et al. 2001), inconsequential interventions that are no longer used (Carey and Carey 1990b), failure to attract patients and families into the intervention (Lehman et al. 1993), treatment drift (Drake et al. 1998), implementation failures (Jerrell and Ridgely 1999), enormous attrition (Hellerstein et al. 1995), or other problems. RCTs that are currently being conducted study specific component interventions (e.g., a type of group or family intervention) but will yield no information on how the component intervention fits into an effective service system. If, as some evidence suggests, the key issue is engagement and retention in any active treatment, studies of interventions to increase participation rather than of refinements of specific component interventions may be more appropriate (Carey and Carey 1990a). Despite all of these problems, several quasi-experimental studies of dual-diagnosis programs have been relatively feasible to implement (Jerrell and Ridgely 1995; Drake et al. 1997; Carmichael et al. 1998; Ho et al. 1999; Kasprów et al. 1999; McHugo et al. 1999; Brunette et al. 2001; Aguilera et al. 2002). Although the interventions in these quasi-experimental studies differ from study to study, the consistent finding across studies is that integrated treatment is more effective than parallel treatment and that several component interventions, such as combining individual, group, and family interventions, are always present in successful programs (Drake et al. 2001). Thus, quasi-experimental studies have been more successful and informative than RCTs in a difficult and emerging area of psychosocial research.

Concluding Remarks

After the complexities of studying psychosocial rehabilitation methods are considered, studies of medication effectiveness can look straightforward. It is, therefore, important to again ask whether, for the question at hand, the design being proposed is the most rigorous approach feasible and whether that approach is likely to yield answers that will move the field ahead. If both these conditions are met and the question being studied is important, then the study likely has merit. We know that pharmacotherapy is a central feature of the treatment of schizophrenia, yet we know that pharmacotherapy alone is not sufficient treatment; hence, the challenge remains to find ways to study the effectiveness of alternative psychosocial interventions. We must continue to expand our understanding of ways to bring stronger designs to important questions for which DBRTs may have limited external validity. In such situations, as we design studies, review proposals, and serve as mentors, the first cut in assessing the quality of the design is not, Is this a DBRT? Rather, the question is, Is this the strongest design possible to address this issue? Designers of the CATIE trial made necessary choices in its trial design that led to trade-offs among controlling extraneous variables, more rigorously evaluating treatment effects and the manner in which treatments are administered. Although the external validity of the trial’s results is potentially affected and the answers it provides may not be final or complete, the CATIE study should yield important information about the effectiveness of treatments in mental health care settings, although the answers may not be definitive.

References


**The Authors**

Susan M. Essock, Ph.D., is Professor and Director of the Division of Health Services Research, Department of Psychiatry, Mount Sinai School of Medicine, NY; and Evaluation Director, Mental Illness Research, Education, and Clinical Center, Veterans Affairs New York Healthcare System, Bronx, NY. Robert E. Drake, M.D., Ph.D., is Professor, Department of Psychiatry, Dartmouth Medical School, Lebanon, NH. Richard G. Frank, Ph.D., and Thomas G. McGuire, Ph.D., are Professors, Department of Health Care Policy, Harvard University, Cambridge, MA.