Enhancing Validity in Co-occurring Disorders Treatment Research

Gregory J. McHugo1,2, Robert E. Drake2, Mary F. Brunette2, Haiyi Xie2, Susan M. Essock3, and Alan I. Green4

2Dartmouth Psychiatric Research Center, 2 Whipple Place, Suite 202, Lebanon, NH 03766; 3Division of Health Services Research, Mt Sinai School of Medicine; 4Department of Psychiatry, Dartmouth Medical School

Despite the high prevalence of co-occurring mental health and substance-use disorders, there has been a relative lack of treatment research with this population, and the existing research often has limited validity. This article explores some of the barriers to the conduct of research on promising interventions for substance-abuse treatment for people with co-occurring disorders, using the concepts of external and ecological validity to make recommendations for future investigation. The central recommendation is to move rapidly from efficacy studies to more credible and valid effectiveness studies, in order to facilitate the adoption of evidence-based interventions in routine practice settings.

Key words: inferential validity/external validity/ecological validity/co-occurring disorders/dual diagnosis/effectiveness research

Introduction

Co-occurring severe mental illness and substance abuse, often called dual diagnosis or dual disorders, is a major public health problem.1 This fact has been well established since the 1980s, yet there has been relatively little research on treatment for people with co-occurring disorders. These patients are unstable,2 diagnostically complex,3 difficult to recruit to studies,4 difficult to engage in treatment,5 and especially difficult to retain in treatment.6,7 These factors lead to the exclusion of patients with co-occurring disorders from controlled research studies and to difficulty completing studies aimed at this population. In addition, the presence of separate centers within National Institutes of Health for the study of alcohol (National Institute on Alcohol Abuse and Alcoholism), drugs (National Institute on Drug Abuse), and mental illness (National Institute of Mental Health) has led to an unfortunate lack of responsibility and cooperation concerning the study of co-occurring disorders.

Treatment research of co-occurring disorders has been summarized recently in terms of specific psychosocial interventions,8 residential interventions,9 treatment principles,10 and pharmacotherapies.11 Although more than 40 controlled studies show advantages for specific interventions, there have been few replications. In many cases, the experimental intervention represents a closer integration of mental health and substance-abuse treatments than the control intervention, but there is little consistency across studies in terms of designs, patients, interventions, and outcome measures.8 Many of the studies are quasi-experimental rather than experimental, different types of patients are included in studies, many of the interventions are complex amalgams, and outcomes and measures vary considerably.10 Thus, after 20 years of research, there remains a lack of strong and clear evidence regarding effective engagement, treatment, and rehabilitation interventions for people with co-occurring disorders. Furthermore, even where the evidence base is consistent, there has been reluctance among service providers to implement new interventions for co-occurring disorders, leading to the well-documented science to service gap.1

We believe that it is time to examine the reasons why there has been a lack of research and adoption of interventions for patients with severe mental illness and a co-occurring substance-use disorder. We propose that progress in the past has been slow because conventional research methods often lead to studies with low validity when applied to this population.

Validity

Based on the work of Campbell and colleagues, validity refers to the truthfulness of inferences drawn from research findings. Inferential validity was originally discussed as internal vs external validity,12 although more recently, internal validity has been expanded to include statistical conclusion validity and external validity has been expanded to include construct validity.13,14 Internal validity pertains to the elimination of bias from the cause-effect relationship. Statistical conclusion validity pertains to the appropriate use of statistics. Construct validity
pertains to the generalizability of the findings across the operations and measures of a study, and external validity pertains to the generalizability of the findings across people, settings, and time. Co-occurring disorders’ research has been weak in all 4 domains of inferential validity. As the evidence-based medicine movement grows within behavioral health, the lack of valid research will hinder the development of evidence-based practices, treatment guidelines and algorithms, and decision supports for co-occurring disorders.

The accrual of evidence in support of an intervention comes from many directions and with various claims to scientific rigor and inferential validity. The challenge before the co-occurring disorders field now is to design intervention studies that address key clinical questions in a manner that is scientifically sound and that generalize to practitioners and patients in routine practice settings. This brings to the fore the tension between efficacy and effectiveness research. The basic issue in balancing efficacy vs effectiveness involves the extent to which the research reflects ideal scientific conditions (high internal validity) vs real-world practice conditions (high external validity).15

The randomized controlled trial (RCT) is the gold standard of efficacy research because it provides the strongest basis for drawing valid causal inferences. That is, by eliminating potential sources of bias, the well-conducted RCT can isolate the intervention as the cause of the outcome. The common approach in most efficacy research is to maximize internal validity by using a narrowly defined population, interventions designed specifically for this narrow population, highly trained researchers, tightly controlled randomized trials, short research follow-ups, and proximal outcome measures. Efficacy research maximizes scientific control and inferential validity concerning the treatment effect, but studies of this type often produce such narrow and non-generalizable findings that they do not apply to typical patients, do not lead to the proposed next stages of intervention development, and fail to influence routine clinical practice.16-18 Moreover, this approach assumes the stage-wise development, validation, and dissemination of interventions (eg, Rounsaville et al19), but this process can be so slow that new interventions displace the tested ones before the tested ones are widely disseminated.

While researchers often test theoretically sound interventions in carefully controlled situations, common real-world interventions—those that are regularly used by clinicians, believed to be effective, and therefore warrant testing—are often difficult to study in RCTs. Examples from the co-occurring disorders field include injectable antipsychotic medications, clozapine, dual-recovery self-help interventions, and long-term residential treatment. Each of these interventions, except clozapine, is widely used in existing dual-diagnosis programs but has not been studied carefully, for different reasons. For example, residential treatment programs must follow counteracting and restrictive regulations of the Department of housing and urban development (HUD), and clinicians are often reluctant to assign unstable patients to clozapine by randomization. Despite several quasi-experimental studies that support residential dual-diagnosis treatment9 and clozapine,11 there have been no successfully completed rigorous RCTs of either intervention.

Despite the power and prestige of the RCT, there are many situations in which true experiments are impossible, unnecessary, premature, or uninformative.20 Even if randomization is possible in a given situation, experimental control may not be, so each threat to inferential validity has to be evaluated carefully. For example, randomization to a community treatment team vs standard case management may go smoothly only to have clients fail to receive the assigned intervention for a wide variety of client-level reasons, such as being jailed, moving out of state, or deciding to discontinue treatment, or system-level reasons, such as a reconfiguration of available services during the course of the study. The longer the treatment and follow-up periods of a study, the more likely such challenges to intervention fidelity and sample retention become. A study with substantial attrition would indicate how many individuals, if offered the treatment under these circumstances, would participate and would highlight the challenges to mounting the intervention in question but would not provide the answer to the basic clinical question regarding effectiveness.

In many cases where an RCT has not been possible, dual-disorders studies have employed a range of quasi-experimental designs to address important research questions often in naturalistic settings.10 These designs are useful for the study of system-level policy changes, the examination of a longer term perspective on outcomes of those situations when feasibility and pilot data are needed, and occasions when there are barriers to randomization. In addition, qualitative research methods can yield deeper understanding of specific issues due to their focus on subjective experience and narrative data. They have been used to study the validity of measures, patient attitudes and motivations, reasons for noncompliance, and system barriers to dissemination and implementation.21 For the study of co-occurring disorders, qualitative methods have been used to understand recovery from substance abuse from the patients’ perspective, to assess natural social support networks, and to explain treatment refusal or dropout.22,23

This discussion highlights the value of the RCT for determining intervention efficacy, but it also reveals the shortcomings of efficacy research for developing the evidence base in support of a given intervention. Other designs and methods may be more practical and informative, and they may yield findings with higher external validity. The various approaches can be viewed as complementary rather than oppositional because evidence of
the efficacy of an intervention is necessary but seldom sufficient to ensure its widespread adoption in routine practice settings. We believe that the widespread adoption of interventions to treat substance abuse among people with co-occurring disorders will occur only when studies lead to inferences with high external validity and include procedures with high ecological validity.

**External Validity**

A study has high external validity if its findings apply to the intended population of patients, providers, and settings. Despite the importance of external validity, the deliberate focus on internal validity among investigators has led to a great deal of irrelevant and nongeneralizable research. Rothwell documented the lack of attention to external validity in research across a wide range of medical conditions and proposed higher standards for both single studies and systematic reviews. Most telling is the oft-repeated claim by providers that the failure to adopt new practices is due to the lack of credibility and applicability of the evidence base.

To illustrate this problem, consider the following hypothetical question, which is based on approximate population estimates. If only 50% of people with schizophrenia and co-occurring substance abuse are in treatment, if only 20% of those in treatment take antipsychotic medications as prescribed, and if only 10% of those who take medications are interested in adding another medication for substance abuse, do we want to study the new medication for the 1% of eligible patients or do we want to study interventions for engagement in long-term medication management for the 99% of eligible patients? This example is not farfetched. By the time inclusion and exclusion criteria have been imposed, and other recruitment procedures have been implemented, the enrolled sample may differ in important ways from the majority of patients who are seen in routine practice. It may be more sensible for clinicians to ignore a study’s results entirely rather than to try to adjust the risks and benefits of a treatment that was evaluated in a patient group that is only remotely similar to theirs.

**Ecological Validity**

In addition to stressing the importance of external validity, we also stress the need for studies with greater ecological validity. Although defined variously (eg, Bronfenbrenner and Brunswick), we use ecological validity here to mean the extent to which the methods, settings, and interventions of an experiment approximate the real-life situation that is under study. Therefore, to enhance ecological validity, the challenge is to conduct studies that mimic the clinical realities of the engagement, treatment, and rehabilitation of patients with co-occurring disorders.

One such reality is the decision-making process in which clinicians and patients engage regularly. If, on the one hand, a study contains a rigid protocol for making decisions about treatment dose and duration without regard to co-occurring conditions, engagement in services, adherence to treatment, and patient preferences, it will have low ecological validity because its procedures may not simulate the complex decision process that is used in routine practice. If, on the other hand, a study protocol reflects the usual decision making that determines which intervention to try first, how long a given trial lasts, what outcomes are tracked, how to sequence interventions, and how to handle nonadherence and dropout, it will more closely mimic the decision process in routine care and have higher ecological validity.

The remainder of this article focuses on ways to enhance both external and ecological validity in dual-disorders research, without compromising internal validity. The primary advantage is that the dissemination and implementation of effective interventions will be facilitated due to the credibility and clinical meaningfulness of the research rather than assumed due to the statistical significance of the findings.

**Design**

Ideally, design strategies reflect current clinical thinking and practice and also include the largest proportion of participants, test the most relevant interventions, and account for patient preferences. Unfortunately, dual-disorders research has been hampered by the unquestioning adoption of RCT methods for effectiveness trials. Alternatives to the conventional RCT have emerged recently, and they may lead to greater external and ecological validity, while maintaining internal validity, in studies of co-occurring disorders.

**Sequential Adaptive Treatment Designs**

In treating chronic relapsing disorders, clinicians make decisions over time that depend on multiple factors, such as the availability of treatments and the allocation, adherence, and response to them. Historically, treatment research for chronic disorders has not evaluated what clinicians actually do. Instead, new treatments are evaluated in isolation relative to standard care, thereby bearing little on actual practice. Fortunately, recent developments have led to the elaboration of experimental designs that permit evaluation of adaptive treatment strategies (eg, Collins et al, Lavorie and Dawson, and Murphy). In cases where treatment guidelines are well specified, fixed adaptive designs are appropriate. These designs contrast algorithmic treatment strategies with each other or with usual care in order to evaluate the effectiveness of adhering to a predefined sequence of decisions concerning treatment. For example, fixed
adaptive designs have been used to evaluate the treatment of depression in primary care settings. In these studies, rule-based but flexible strategies for care management, which included education, medication, specialty care, and adherence, side effect, and outcome monitoring, were contrasted with usual care and found to be more effective.

Because treatment for co-occurring disorders does not have well-established guidelines currently, randomized adaptive designs are an appealing option. By adding random assignment at each decision point, studies can contrast multiple treatment strategies, each of which involves a sequence of options that depend on factors such as response to prior treatment, adherence, and intervention modality. These studies are efficient in developing and refining sequential adaptive treatment strategies, although they require large sample sizes. Ideally, each effective sequence is then codified as a treatment algorithm and contrasted with standard treatment in a conventional 2-group RCT. As examples, randomized adaptive designs have been used in recent trials to evaluate treatment for depression, Alzheimer disease, alcohol addiction, and cocaine addiction. For the study of treatment of alcohol abuse among people with co-occurring disorders, such a design might include an initial period of dual-diagnosis case management, after which nonresponders are assigned randomly to next-level interventions (eg, a medication strategy or a dual-diagnosis group intervention). Continued nonresponse at the next assessment point could lead to a switching or augmenting of interventions, with or without randomization. Responders at any level of the design could be randomly assigned to one of several relapse prevention interventions. With sufficient sample size, such a design could rapidly test multiple interventions and could determine the potency of various sequences of interventions in relation to the long-term outcome of sustained abstinence. By evaluating long-term outcomes and allowing an informed sequence of treatment decisions, these designs have greater ecological validity and could move the field more quickly toward useful guidelines for the timing, matching, and sequencing of treatments.

Equipoise-Stratified Randomization
Randomized trials of multiple interventions often are limited by the size and representativeness of the enrolled sample because each participant must be willing to accept all possible treatment options in order to be assigned randomly to any one of them. If a prospective participant rejects (eg, refuses medication), or is not appropriate for (eg, has no family), a specific intervention, that participant cannot be included in the study. One proposed solution to this problem is equipoise-stratified randomization. Equipoise means that prospective study participants (patients, clinicians, or both) regard a set of treatment options as essentially equivalent in terms of the likelihood of success. Equipoise-stratified randomization is a method for retaining the maximum number of participants possible, while still allowing contrasts between conditions (interventions) to which participants have been assigned randomly. At the point of consent, participants are informed of the treatment options offered in the experiment, and they decide which options are acceptable to them. As participants enroll in the experiment, a number of strata will emerge depending on the subsets of the treatment options that participants are willing (and eligible) to accept. The only people who are excluded are those who are willing to accept only one or none of the available options. Otherwise, the study can accommodate all volunteers willing to be assigned to at least 2 of the treatments under study, thereby enhancing external validity. More work needs to be done concerning the analysis and interpretation of studies using equipoise-stratified randomization, but the emergence of this approach underscores the need for solutions to problems of recruitment, which often result in irrelevant and ignorable findings.

These innovations in experimental design can greatly enhance the external and ecological validity of research concerning co-occurring disorders. Studies that examine the timing and sequencing of psychosocial, medication, and residential interventions are sorely needed. Factors such as intervention cost and burden could be taken into account in the sequencing, and the treatment decision process could more closely approximate that used in routine settings. The addition of equipoise-stratified randomization could enable the findings to generalize to a much broader segment of the dual-diagnosis population.

Length of Follow-up
Long-term outcomes should be the goal for research that seeks to understand chronic, fluctuating illnesses. Brief interventions and short-term follow-ups may test short-term compliance rather than sustained remission and personal recovery. For example, studies of injectable antipsychotic medications tend to show short-term gains but no long-term effects. Long-term treatment outcome studies are expensive and difficult to conduct, and yet, the answers to many of the most important questions facing the field can only be answered by such studies. If we are to bridge the gap between efficacy and effectiveness studies, one solution will be to pay more attention to identifying short-term outcomes that predict long-term recovery.

When studies of co-occurring disorders are designed to evaluate adaptive treatment sequences, the natural focus will be on long-term outcomes. An effectiveness study to evaluate medication options to treat alcohol abuse may take a year to conduct, with another year of follow-up, and therefore, an end-point outcome might be whether or not participants attain remission or sustained abstinence.
A typical efficacy trial of one of the medications might last for 3–6 months, and the outcome might be the change in the average daily quantity of alcohol consumed. Surely, the long-term perspective of the former study leads to outcomes that are more meaningful and clinically significant for patients and clinicians, although short-term studies to establish treatment efficacy and tolerability are needed first. One implication of this is that efficacy studies should include short-term (proximal) outcomes that predict long-term (distal) outcomes rather than solely using short-term outcomes that are supposed surrogates of the long-term outcomes but may not actually predict them.

Recovery from both mental illness and substance abuse is a longitudinal process, not a short-term outcome. On the one hand, long-term illnesses need to be studied in relation to long-term outcomes because the short-term course involves fluctuations and brief changes that may not predict eventual recovery. Everything we know about dual disorders implies a long-term course of recovery. On the other hand, system-level outcomes involve improving the processes of care. Measuring such relatively short-term events as changes in the likelihood of clients receiving particular evidence-based interventions (processes of care) is a means of characterizing system-level outcomes. To the extent that long-term recovery depends on short-term outcomes (eg, treatment access and adherence), research must take both into account.

**Outcomes**

Current ethical, clinical, and research perspectives mandate that we study outcomes that are meaningful to both clinicians and patients. Patients value goals that are embodied in the concept of recovery, which is defined by the President’s Commission on Mental Health as “living, learning, working, and participating fully in one’s community”. In focus groups, ethnographies, and research interviews with dual-diagnosis patients, we have found that they also identify several other goals: avoiding negative health outcomes, like HIV and hepatitis C infection; avoiding aversive living states, like homelessness and hospitalization; avoiding criminalization and incarceration; managing their own illnesses, including mental illness and substance abuse; avoiding negative side effects related to medications; and avoiding victimization, stigma, and interpersonal abuse. Sometimes, they identify goals that are qualitatively different from the outcomes that are typically assessed by researchers. For example, dual-diagnosis patients often identify having friends and intimate partners who are not substance abusers as an important component of their recovery. Likewise, for people with co-occurring disorders, treatment discontinuation at any stage of recovery is associated with negative outcomes and should be considered an important outcome in its own right. Too often, treatment nonadherence and dropout are considered nuisances rather than key outcomes, although they are among the preconditions that determine the magnitude and scope of the treatment effect.

Deciding what outcomes to measure is a central issue when studying co-occurring disorders. Mental health and substance-abuse outcomes are relatively independent dimensions, and they are largely independent of outcomes in other domains, such as employment, quality of life, general health, and residential status. Having used a multidimensional assessment, the challenge then is to decide whether to treat the outcome within each domain separately and hierarchically (ie, primary vs secondary outcomes) or to combine outcomes into a single indicator of a complex construct like recovery (eg, Drake et al and Xie et al). One proposed way to circumvent this decision is to use an analytic approach that selects the outcome of greatest change for each participant prior to testing the contrast between groups. This method acknowledges that multiple outcomes may be affected by an intervention and that change may happen on different outcomes for different people.

A further consideration concerning outcomes is interpreting their clinical significance, not just their statistical significance. Continuous measures have advantages for statistical analysis, but they do not readily translate into clinically meaningful conclusions. Proposed solutions to this problem include comparison with established norms or use of the reliable change index. As an alternative to statistical solutions, investigators often define outcomes a priori in clinically meaningful ways. By doing so, a statistically significant difference can be evaluated more readily for its clinical significance.

Often the immediate questions following demonstration of a treatment effect are for whom and how does it work? Moderator and mediator effects, respectively, take direct aim at these questions. A priori specification of a model of the treatment and recovery process is important when deciding what to measure as possible moderators or mediators. Recent discussions by methodologists have improved the understanding and analysis of moderator and mediator effects in longitudinal outcome studies and clinical trials.

Moderators are often attributes of the study participants that influence the response to treatment. Moderator effects are most often detected statistically as 3-way interactions in longitudinal data analysis (Treatment Group by Moderator Level by Time). Because they are observed characteristics of study participants, moderator effects are necessarily correlational and must be interpreted accordingly. For example, among patients with co-occurring disorders, antisocial personality disorder or the presence of severe physiological dependence may moderate the impact of treatment for alcohol abuse, but causal mechanisms cannot be specified.
Mediators are links in a causal chain of outcomes, whereby levels of an intervening (proximal) outcome causally influence more distal outcomes. Mediators are often process variables, such as services received or knowledge gained, which occur after enrollment in an intervention but before outcomes are assessed. For example, the effect of naltrexone on alcohol abuse may depend on adherence with the medication regimen. In this case, treatment adherence mediates the effect of naltrexone on alcohol use. A range of statistical methods has been proposed to detect mediator effects, and they differ in the trade-off between Type I errors and statistical power.50

Measuring Substance Use

Whether the primary outcome or not, substance use will be part of any study of an intervention for co-occurring disorders. Unfortunately, there are limitations on the reliability and validity of measures of substance use from any source, and the target behavior is not easily measured over extended periods of time. Moreover, there are numerous possible sources of bias in verbal reports in general, many of which are likely when reporting illegal and socially undesirable behavior like drug use. There are also limitations with collateral reports and biological measures.

Currently, the most reliable and valid way to assess substance use is to gather assessments from multiple sources, using multiple methods, and to bring those measures into a rule-based process for developing consensus ratings. Specifically, we have used participant self-report, clinician ratings, and biological indicators to triangulate on a reliable and valid assessment of substance-use disorder (eg, Drake et al51 and Essock et al52). Given the uncertainty surrounding any single measure of substance use, a standardized process for combining information from multiple sources provides the most valid measure possible, especially when the outcome of interest is clinically meaningful, such as presence or absence of stable abstinence, the status of alcohol-use disorder, or the days of drug use in the past 6 months.

Interventions

Because we are interested in interventions that improve outcomes of usual care, clarifying and documenting usual care is critical. Based on the current state of the evidence, what is ethical and evidence-based to include in usual care for patients with co-occurring disorders—clinical case management, cognitive behavioral therapy, referral to self-help, or illness self-management? Unfortunately, the evidence is not yet strong enough for numerous specific dual-disorders interventions to make this decision. Yet, it is crucial to standardize usual care in order to improve the inferences drawn from studies of co-occurring disorders. Many clinics and other settings are already providing some form of dual-diagnosis treatment; the challenge is to bring it under algorithmic control via explicit protocols so that interventions are tested against meaningful standards.

Even if usual care can be standardized prior to the start of a research project, it may be difficult to keep it standardized over time. For example, randomization within clinics to either usual care or a new intervention can lead to numerous threats to inferential validity. It is difficult to control drift, diffusion, and compensatory efforts when control group (ie, usual care) clinicians become aware of experimental interventions. In addition, various efforts to rescue clients, such as transferring them to supervised housing or changing medications, may also, of necessity, undermine experimental control. Quasi-experimental research comparing clinics (eg, one provides usual care and one provides usual care plus the intervention) may be more feasible, although some of these same threats to inferential validity apply to them as well, and the lack of randomization leaves inferences vulnerable to selection bias. Even in the most rigorously designed studies, numerous trade-offs must be evaluated because randomization can cause, as well as overcome, threats to validity.

Given high uncertainty, due to the lack of direct evidence, how should interventions be selected? It may make sense to start with interventions that have shown efficacy in small trials in the dual-disorders population or in related populations. At this time, motivational interviewing linked with cognitive behavioral therapy, contingency management, residential treatment, and special medications for substance abuse (eg, naltrexone and acamprosate) are good candidates.8,11 They have shown promise in the primary substance-abuse treatment population and are ripe for testing in the dual-disorders population.

When considering interventions, adherence to treatment arises as a critical issue. Individuals with dual disorders are notoriously difficult to engage in treatment, for a host of reasons, and therefore, study protocols must anticipate adherence problems. Studies may increase adherence through design features, monitoring protocols, and specific interventions such as contingency management to increase participation.53 The goal is to preserve the integrity of the experiment following randomization, but artificially enhanced adherence comes with the downside of compromising external validity if adherence would be different without the intervention provided by the research protocol. Hence, interventions should not include provisions to maximize adherence that would be infeasible in routine practice settings.

Medications

Patients’ involvement with medication treatment for substance abuse can be conceptualized within a stagewise model. Attitudes about medication are affected by lack
of acknowledgment of a mental illness or denial and minimization regarding substance-use disorder, either of which can lead to disinterest in treatment, in general, and medications, in particular, in the early stages of treatment. As patients learn more about their disorders and become interested in managing them, they are often more willing to try medications as a management tool. Consequently, studies of medications for people with dual disorders are more feasible if they recruit people who are already in either the active treatment or the relapse prevention stages of recovery. Restricting the treated population in this manner limits the generalizability of the study, but paradoxically, it may enhance its external validity, by allocating treatment in the research as it would be allocated in routine practice.

Medications effective for alcohol-use disorders in the general population, such as acamprosate and naltrexone, have modest effects. Consequently, studies of these medications need large sample sizes in order to detect statistically significant differences between groups, and researchers need to be clear about what magnitude of difference is clinically significant. Studies of medications are usually short term and involve measures of substance use that are sensitive to immediate and modest change. These studies do not provide sorely needed evidence concerning the persistence of the changes and the effects of long-term use of the medications.

Studies of medications often incorporate psychosocial interventions to enhance adherence. The question of which psychosocial interventions to study along with medication interventions brings up issues regarding combination treatments vs solo treatments. If the effect of the medication is modest and concomitant psychosocial interventions are usually offered in real-world settings, a more externally valid design is to study the combination of medication plus psychosocial treatment. Placebo plus psychosocial treatment controls could help to assess what portion of improvement is due to the medication component.

Some medications, such as clozapine, may treat psychiatric and substance-use disorders together, whereas other medications are designed to treat the substance-use disorder only and must be added to a regimen that treats the mental illness. The advantage of a single medication approach includes the use of fewer pills, which could be more cost effective, less burdensome, and possibly more appealing to participants, resulting in better adherence. The disadvantages include the requirement to change the primary psychiatric medication and the inability to adjust medication regimens based on symptoms of each disorder.

Because patients with dual disorders tend to experience high levels of psychosocial instability and are often poorly engaged into any sort of services, it is difficult to recruit and retain them in medication studies. Techniques for tracking this population (eg, maintaining multiple contacts from the patient’s social world such as landlords, friends, coworkers, and family) work well for longitudinal studies, but they do not ensure compliance with regular and frequent research appointments to monitor medication treatment. Instead, frequent reminders, transportation, child-care support, and community outreach may be needed in order to enhance patients’ ability to participate regularly in this type of study. A further complication is that patients in later stages of treatment who are more stable are often reluctant to participate in medication studies. Clinicians may view their stability as tenuous, and there may be reluctance on their part to stress the patient by involvement in research or by changing medication. Studies of add-on medications, such as naltrexone for alcohol-use disorders, may be easier to implement because a switch of the medication treating the mental illness is not required for the study.

This discussion underscores the value of, and need for, both efficacy and effectiveness studies of medications for people with co-occurring disorders. Efficacy studies need blinding, placebo control groups, and other procedures of RCTs in order to isolate the treatment effect from other influences. Once a medication effect has been found under ideal conditions, the focus should change rapidly to real-world effectiveness trials where the concern is with such issues as acceptance and adherence, timing and sequencing, clinically meaningful outcomes, persistence of changes over time, long-term side effects, and medication interactions. Findings from both types of studies are needed before evidence-based treatment guidelines and algorithms can be developed.

Practitioners

A difficult trade-off exists when deciding who should provide specialized psychosocial interventions. On the one hand, it may not be best to use highly trained research clinicians because their training, experience, and skills do not generalize to clinicians in routine settings. On the other hand, it is desirable to study a competent implementation of each intervention in order to have a valid test of effectiveness. A reasonable compromise may be to train clinicians in routine settings to a standard of competence and to monitor their practice throughout the study, both clinically (supervision) and empirically (treatment adherence and model fidelity). This has been done in effectiveness studies of integrated treatment for individuals with co-occurring disorders with good success.

Settings

Patients with co-occurring disorders involving severe mental illness often receive services in community mental health centers, hospitals, jails, and homeless shelters, if at all. Studies in routine settings such as these have a good chance of recruiting people who are in different
stages of recovery and are more representative of the population.

One problem with conducting studies in routine settings is that usual care varies widely from setting to setting. As discussed above, there is a pressing need to standardize usual care in order to clarify the effects of new medication and psychosocial interventions. In addition to clarifying usual care, critical features of routine settings can be expected, based on empirical studies, to influence outcomes. For example, the local criminal justice system, the types and amounts of housing programs that exist, and the amounts of family contact and support that patients have may moderate treatment effects. In addition, medications for mental and physical health may interact with treatment for substance abuse and must be monitored closely. Settings will also differ in the extent to which they assess intervention fidelity and assure adherence with the treatments, whether psychosocial or pharmacological.

Patients

A fundamental problem in studies of interventions for co-occurring disorders is the heterogeneity of the patient population. People with co-occurring disorders are diverse, not only in socioeconomic features, personal assets, and community supports but also in clinical and comorbid features. Mental health and substance-use disorders can range widely in type and severity. Patients are also affected by a range of common co-occurring conditions, such as antisocial personality disorder, posttraumatic stress disorder, traumatic brain injury, neurocognitive problems, hepatitis C, and chronic medical problems (eg, diabetes, obesity, and cardiovascular disease). Restricting the population to those with severe mental illness may have modest impact on reducing heterogeneity, although it does increase external validity, because these are the patients who are treated in the public mental health system.

We have already discussed how the typical solution to this problem—a narrow focus on one specific clinical group—often results in studies that severely limit external validity and clinical applicability because they pertain to so few people. But how can we consider heterogeneity in a way that enhances clinically relevant research? We suggest looking for more natural points of cleavage in the population—those that correspond to research findings and clinical realities.

Naturalistic follow-ups, as well as intervention studies, can identify meaningful subgroups of patients with co-occurring disorders. For example, because psychiatric diagnosis among the severe mental illnesses has little or no predictive validity concerning long-term recovery from substance abuse, it may be a poor selection criterion for studies, unless there is a diagnosis-specific intervention. Rather, longitudinal research on substance-abuse recovery suggests that there are distinct subgroups within the larger population of patients with co-occurring disorders, which are characterized in part by their relationships and responses to treatment over time. One group rejects community-based treatments and has uniformly poor outcomes over at least 10 years. A second group engages in treatment rapidly, enters substance-abuse recovery rapidly, and has a stable course over 10 years. The majority of patients occupy 2 groups that have intermediate courses, one with a fluctuating course over 10 years, and another with a course of slow but steady improvements over 10 years. If these subgroups can be validated across samples and characterized by common individual differences and clinical features, it would make sense to classify patients early in treatment for the sake of efficiently providing services that they need, avoiding services to which they may not respond, and designing separate studies for each group.

In addition, studies can be designed to reflect clinical decision making and shared decision making. As described above, adaptive treatment strategies enable the evaluation of care for chronic relapsing disorders in a more ecologically valid fashion. The goal is to make treatment decisions that take the patient’s history, current status, and preferences into account and then to monitor adherence and outcomes in order to adapt appropriately. For example, research on behavior change indicates that treatment is more effective when clinicians and patients think about recovery from substance-use disorders in steps or stages. Based on an approach described by Osher and Kofoed, the patient must first be “engaged” in a therapeutic relationship, defined at minimum by regular participation in treatment. As an example of a study at this stage, Corrigan and colleagues compared motivational interviewing, financial incentives, and barrier reduction in terms of their effect on regular treatment participation at 6 months for clients with substance abuse and traumatic brain injury. They showed that both financial incentives and barrier reduction were superior to motivational interviewing and an attention control.

Following engagement in treatment, many patients with co-occurring disorders remain “unmotivated” to manage their illnesses (persuasion phase). That is, they attend treatment sessions but deny problems with substance abuse, mental illness, or both, and they do not take responsibility for managing these illnesses by using active strategies. The common intervention at this stage is motivational counseling of some type, but other strategies, such as contingency management, family psychoeducation, supported employment, long-acting medications, and mandated treatment, could be tested. Regardless of approach, the goal is to increase the behavioral manifestations of illness management.

Once patients are actively working to acquire the skills and supports for managing their illnesses, they are in the “active treatment” stage. At this stage, a variety of
pharmacological, cognitive behavioral, social network-based, self-help, and other strategies could be tested. When patients are securely in remission (eg, at least 6 months without a relapse of either disorder), they are in the “relapse prevention” stage. A variety of relapse prevention interventions have been developed for mental illnesses and for substance-use disorders, but none has been tested specifically with dual-diagnosis patients. Because patients at this stage of recovery are participating in treatment reliably, are relatively stable, and yet relapse is common, this would be an ideal group for RCTs to examine interventions for successful relapse prevention.

**Summary**

There are fundamental differences between efficacy and effectiveness research, and we have reviewed a number of ways to enhance the validity of the latter, with special emphasis on external and ecological validity. Efficacy studies are necessary in identifying promising interventions that yield a discernible treatment effect under ideal conditions, but their success does not guarantee adoption of the tested interventions. That is, in an efficacy study, under conditions where everything else is equal, a treatment effect can be detected and attributed conclusively to an intervention, but until the intervention has been tested under routine conditions, its actual value remains unknown. Therefore, efficacious interventions need to be investigated more rapidly and extensively under conditions that enhance external and ecological validity. To this end, effectiveness studies can include heterogeneous patient samples, preference-based decisions, adherence interventions and monitoring, routine clinicians and clinical decision making, stage-appropriate interventions, standardized usual care in routine settings, and long-term outcomes that are valued by patients themselves. Effectiveness studies can use innovative designs and procedures to simulate routine clinical processes and to recruit more representative patients, clinicians, and settings.

Abelson included generality and credibility among the 5 criteria that he proposed to evaluate the merits of an experiment. Even if an experiment shines in all other ways, if it is not credible in its methods and generalizable in its findings, it will neither have an impact scientifically nor influence practice. By paying more attention to external and ecological validity, effectiveness studies of treatment for co-occurring disorders will gain easier acceptance among patients, providers, and policy makers, and will be more likely to bridge the gap between science and practice.

**Acknowledgments**

This work was supported by grants from the National Institute on Alcohol Abuse and Alcoholism, National Institute on Drug Abuse, National Institute of Mental Health, and the West Family Foundation. An earlier version of this article was presented at the National Institute on Alcohol Abuse and Alcoholism/National Institute on Drug Abuse/National Institute of Mental Health workshop on “Methodology of Conducting Pharmacologic Clinical Trials in Patients with Alcohol/Drug Dependence and Psychiatric Comorbidity,” Bethesda, MD, in February 2006.

**References**


