Multi-site effectiveness trials of treatments for substance abuse and co-occurring problems: Have we chosen the best designs?

Edward V. Nunes, M.D.[a], **Samuel Ball, Ph.D.**[b], **Robert Booth, Ph.D.**[c], **Gregory Brigham, Ph.D.**[e], **Donald A. Calsyn, Ph.D.**[f], **Kathleen Carroll, Ph.D.**[b], **Daniel J. Feaster, Ph.D.**[g], **Denise Hien, Ph.D.**[a,h], **Robert L. Hubbard, Ph.D., M.B.A.**[q], **Walter Ling, M.D.**[p], **Nancy M. Petry, Ph.D.**[i], **John Rotrosen, M.D.**[j], **Jeffrey Selzer, M.D.**[a,k], **Maxine Stitzer, Ph.D.**[l], **Susan Tross, Ph.D.**[a], **Paul Wakim, Ph.D.**[m], **Theresa Winhusen, Ph.D.**[n], and **George Woody, M.D.**[o]

[a]Columbia University Department of Psychiatry, New York State Psychiatric Institute; National Drug Abuse Clinical Trials Network, Long Island Node

[b]The APT Foundation, Yale University School of Medicine; National Drug Abuse Clinical Trials Network, New England Node

[c]University of Colorado at Denver and Health Sciences Center

[e]Research Institute, Maryhaven, Inc.; National Drug Abuse Clinical Trials Network, Ohio Valley Node

[f]Department of Psychiatry and Behavioral Sciences, Alcohol and Drug Abuse Institute, University of Washington; National Drug Abuse Clinical Trials Network, Pacific Northwest Node

[g]Department of Epidemiology and Public Health, Center for Family Studies, University of Miami Miller School of Medicine; National Drug Abuse Clinical Trials Network, Florida Node

[h]City University of New York; National Drug Abuse Clinical Trials Network, Long Island Node

[q]Duke University Medical Center; Institute for Community-Based Research, National Development and Research Institutes; North Carolina Node

[p]David Geffen School of Medicine at University of California, Los Angeles; National Drug Abuse Clinical Trials Network, Pacific Node

[i]Calhoun Cardiology Center, Department of Medicine, University of Connecticut Health Center; National Drug Abuse Clinical Trials Network, New England Node

[j]New York University School of Medicine, Veterans Affairs New York Harbor Healthcare System; National Drug Abuse Clinical Trials Network, New York Node

[k]Committee for Physician Health, Albany, NY; National Drug Abuse Clinical Trials Network, Long Island Node

[l]John Hopkins University School of Medicine; Bayview Medical Center; National Drug Abuse Clinical Trials Network, Mid-Atlantic Node

Corresponding author: Edward V. Nunes, M.D., New York State Psychiatric Institute, 1051 Riverside Drive, Unit 51, New York, New York 10032; phone: 212-543-5581; fax: 212-543-5326; nunesed@pi.cpmc.columbia.edu .

[m]Center for the Clinical Trials Network, National Institute of Drug Abuse, National Institutes of Health

[n]University of Cincinnati/CinARC; National Drug Abuse Clinical Trials Network, Ohio Valley Node

[o]University of Pennsylvania School of Medicine; National Drug Abuse Clinical Trials Network, Delaware Valley Node

Abstract

Multi-site effectiveness trials such as those carried out in the National Drug Abuse Treatment Clinical Trials Network (CTN) are a critical step in the development and dissemination of evidence-based treatments, because they address how such treatments perform in real-world clinical settings. As Brigham and colleagues summarized in a recent article (Brigham, Feaster, Wakim, & Dempsey, 2009), several possible experimental designs may be chosen for such effectiveness trials. These include: 1) A new treatment intervention (Tx) is compared to an existing mode of community based treatment as usual (TAU): Tx versus TAU; 2) A new intervention is added to TAU and compared to TAU alone: Tx + TAU versus TAU; or 3) A new intervention is added to TAU and compared to a control condition added to TAU: Tx + TAU versus control + TAU. Each of these designs addresses a different question and has different potential strengths and weaknesses. As of December 2009, the primary outcome paper had been published for 16 of the multi-site randomized clinical trials conducted in the CTN, testing various treatments for drug abuse, HIV risk behavior, or related problems. This paper systematically examines, for each of the completed trials, the experimental design type chosen and its original rationale, the main findings of the trial, and the strengths and weaknesses of the design in hindsight. Based on this review, recommendations are generated to inform the design of future effectiveness trials on treatments for substance abuse, HIV risk, and other behavioral health problems.

Introduction

The National Drug Abuse Treatment Clinical Trials Network (CTN) was founded with the mission to "transform the treatment of addiction in this Nation with science as the vehicle" (paraphrasing Alan Leshner, NIDA Director at the inception of CTN). The science is to conduct multi-site randomized clinical trials to test the effectiveness of promising new treatments in real world, community-based treatment settings—Stage III research, according to the well accepted stage model of treatment development (Rounsaville, Carroll, & Onken, 2001). Such trials address the ultimate questions for a new treatment, regarding how the treatment works when applied in the real-world treatment system--in community-based treatment settings and in the hands of community-based practitioners (Institute of Medicine, 1998).

As the CTN began to launch trials, it became clear that the science of effectiveness trials presented unique challenges, calling for experimental designs distinct from classic Stage II efficacy trials (Brigham, Feaster, Wakim, and Dempsey, 2009). A classic efficacy trial is designed to maximize internal validity and generally randomly assigns carefully selected patients to a well-specified new intervention or treatment (Tx), or an equally well-specified control condition (Control), usually controlling for professional attention and other non-specific elements of treatment. The background setting or platform for the trial, usually a research clinic (or clinics in the case of multisite trials), is also carefully controlled and often artificial in comparison to community-based treatment settings. In contrast, effectiveness trials seek to address external validity, and generally seek to recruit representative samples of patients from the community, and to address questions of how a new treatment intervention performs in relation to Treatment as Usual (TAU) in community settings.

This paper reviews the research designs chosen for 16 multi-site randomized clinical trials from the CTN for which the primary outcome paper was published as of December 2009 and reviews the designs chosen for each of the trials in relation to the experience conducting them and the main findings. We begin with a summary of each of the design types and their theoretical strengths and weaknesses. We then catalogue each of the published CTN trials according to which design type it represents and examine for each trial how the chosen design performed in light of the original goals and the actual findings. The ultimate goal is to derive lessons to inform the optimal design of future trials in the addictions and related fields.

Overview of Design Options for Randomized Effectiveness Trials

Brigham, Feaster, Wakim, and Dempsey (2009) recently published a review of four prototypical effectiveness designs, drawing from their experience considering design options during the development of several CTN trials. Only 2-group designs were considered where patients are randomized to some new intervention versus a control condition. These designs (see Table 1) were placed in order from strongest generalizability and external validity (Designs 1 and 2) to strongest internal validity (Designs 3 and 4).

Design 1

Participants are randomized to a New Intervention (Tx) versus treatment as usual (TAU), such that Tx substitutes entirely for TAU (Tx vs TAU).

Design 2a

Participants are randomized to a New Intervention added to TAU, (Tx + TAU vs TAU; New intervention is an add-on to existing TAU).

Design 2b

Participants are randomized to a New Intervention that substitutes for a component of TAU, (Tx + TAU vs TAU; New intervention substitutes for part of TAU).

Design 3

Participants are randomized to a New Intervention added to TAU, versus some specified Control intervention added to TAU (Tx + TAU vs Control + TAU).

Design 4

Participants are randomized to a New Intervention versus a standardized control intervention (Tx vs Control), this being essentially a variant on the classic Stage II efficacy trial.

Note that Design 2 in the original conceptualization (Brigham, Feaster, Wakim, and Dempsey, 2009) has been subdivided into Design 2a and Design 2b, depending on whether the new intervention is an add-on to TAU, or substitutes for some existing component of TAU.

Each design was originally characterized according to 12 methodological strengths and 14 weaknesses (Brigham, Feaster, Wakim, and Dempsey, 2009). For this paper, we synthesized these strengths and weaknesses into 6 methodological issues on the basis of which we analyze the designs and how they performed in each of the CTN trials. Namely, each design addresses different questions and has different strengths and weaknesses with respect to external and internal validity, effect size and power to detect intervention effects, and the variability between sites in multi-site trials.

1) Primary question addressed—Each design addresses a different question, and it is important that the question chosen be the most useful for informing clinical practice, treatment programs, and treatment systems. Design 2a addresses the impact of adding a new intervention, such as Prize Incentives (Petry, 2000; CTN 0006: Petry et al., 2005a; CTN 0007: Peirce et al., 2006), to existing treatment at community-based treatment programs (TAU). Design 2b addresses the impact of substituting a new intervention for some existing component of TAU, such as substituting Motivational Interviewing for the usual intake interview (CTN 0004: Carroll et al., 2006). Design 3 addresses the impact of adding or substituting a new intervention onto/into TAU, over and above the effect of a specified control condition. The control condition may be intended to systematize some aspect of TAU and reduce variability between sites (e.g., clonidine as a standard non-narcotic detoxification regimen in trials of buprenorphine for opioid detoxification—CTN 0001 and CTN 0002 (Ling et al., 2005)) or to control strictly for attention and other nonspecific elements of treatment (e.g., Women's Health Education as control for Seeking Safety, a cognitive behavioral intervention for substance dependent women with PTSD—CTN 0015 (Hien et al., 2009).

The choice among designs, and their primary questions, should be driven, in part, by the state of the evidence on a given intervention from prior efficacy trials and usual clinical practice. When the evidence for efficacy is strong, there is less concern about controlling for non-specific elements. When the evidence for efficacy is weaker, or when an intervention has been substantially modified to make it suitable for community-based treatment, then the field needs to know whether the specific elements of the new intervention that were deleted or retained contribute to efficacy.

2) Suitable treatments to test—Designs 1 and 4 are suitable for interventions that would substitute entirely for usual community-based treatment (TAU) if found to be efficacious. Designs 2a, 2b, or 3 are suitable for new interventions that would be either added to TAU, or substituted for some existing component of TAU. Only one CTN trial has used Design 1, a study of Brief Strategic Family Therapy as an alternative to usual aftercare among adolescents discharged from residential treatment (Robbins et al., 2009), although the primary outcome analysis is not yet published. No CTN trials have as yet used Design 4.

3) External validity—External validity is achieved to some degree by recruiting representative samples from community-based treatment settings, although the design itself is also germane. By addressing how a new intervention performs in comparison to current practice (TAU), Designs 1, 2a, and 2b address questions of most direct relevance to providers, program directors, policy makers, and payers. Namely, "should we add or switch to the new intervention?" And, "what better outcome can be expected compared to current practice?" By inserting the specific control condition, which is to some degree artificial and not part of TAU, Design 3 cannot directly answer those questions. To the extent that the control condition matches the new intervention on quantity and quality of attention, Design 3 addresses the extent to which the specific elements of the new intervention are needed to improve outcome. This may be important information for providers and payers. New interventions may be more expensive to implement due to cost (e.g. prize incentives, or a medication such as buprenorphine) and the staffing needed to administer them, or more difficult or time consuming to train and supervise (e.g. a psychotherapeutic technique such as Motivational Interviewing).

4) Internal validity—Threats to internal validity occur when the intervention effect measured in a study, compared to the control condition, may reflect factors not specific to the intervention. Much of the threat to internal validity is handled through random assignment. However, in the effectiveness designs, patients may get more time and attention

from staff in the new intervention than in the control condition. In addition to the quantity of attention, the quality of attention may also have an effect. The new intervention may be delivered with more general skill, engendering better treatment alliance, or more enthusiasm, creating positive expectations. Or, the new intervention may have a good reputation among patients.

Negative expectations may also surround a control condition. Substance dependent patients tend to have strong opinions about treatments, good and bad. Patients may be disappointed or demoralized if assigned to the control condition. The latter was originally dubbed the "Wait-list control effect" in psychotherapy designs where the control condition was, literally, assignment to a waiting list. Such negative expectancy effects (we will refer to them from here forward as "Wait list control" effects), are of particular concern, because, like positive attention effects, they also bias a study in favor of detecting a treatment effect, but in this case it is truly an artifact of the study itself--The control condition performs worse than it normally would because of the study, and the new intervention may look good in comparison even without having any inherent effectiveness. An indicator of this problem would be unusually bad outcome, or high dropout rate in the control condition.

The specified control conditions in Designs 3 and 4 protect against these potentially biasing effects of attention and expectancy, to the extent that the controls actually match the new intervention in quantity and quality of attention. In CTN studies using Design 3, some control conditions were only intended to standardize a minimal intervention that is usually part of TAU, and did not match the intervention condition in time. Examples include the 1-session HIV education control conditions that were used in CTN 0018 (Calsyn et al., 2009) and CTN 0019 (Tross et al., 2008) that tested 5-session skills-oriented HIV risk-reduction interventions. Designs 1, 2a, and 2b are more vulnerable to these biasing effects. In Design 2a, where a new intervention is added to TAU, and compared to TAU alone, patients in the new intervention get more time and attention from staff. Design 2b, where the new intervention substitutes for part of TAU, the amount of time and attention is equated, removing that threat to internal validity, although there could still be positive expectancy effects or Wait-list effects. The only way to measure the extent of such attention effects in designs 2a or 2b would be to include a second control group in the design that controls for attention, in addition to the TAU control—a hybrid of designs 2 and 3. Only one CTN study, so far, used a 3-group design, which can be viewed as fitting this model (Campbell et al., 2009).

5) Effect size and power to detect intervention effects—The size of the treatment effect is ultimately of great interest to providers and payers. A given treatment might achieve statistically superior outcome, but the size of the effect is so small as to be judged not worth the effort or expense. Designs 1, 2a, and 2b may be more likely to be associated with larger treatment effects, since the effect estimated is the sum of both specific effects of the new intervention and non-specific effects such as those related to attention. This summed effect will often be what is most interest to providers and payors. To the extent that the specified control conditions in Designs 3 and 4 match on quantity and quality of attention, Designs 3 and 4 estimate only the specific effects of the new intervention, and thus may yield smaller effect sizes. A caveat is that larger effect sizes are also of interest to researchers because they mean the study is more likely to be perceived as a success and published in high impact journals. This creates a potential for bias in the mind of the researcher regarding the choice of design, which may or may not correspond with the design that answers the questions most needed by the field.

6) Effects of site variability—In a multi-site trial, the effectiveness of the TAU component of the control condition is likely to vary from one community-based program to

another--<u>site effects</u>, or effectiveness of the new intervention being tested may vary between sites--<u>site by treatment interaction</u>. Effectiveness of TAU may vary between community-based treatment programs due to differences in severity or treatment-resistance of their patient populations, differences in the skill level of clinical staff, or differences in the services provided or the overall program atmosphere. Or, a new intervention may work better in the context of TAU at some programs than in others. Such effects add variability and may erode power but also address important questions about variations in TAU and the context of TAU under which the new intervention is more or less likely to be effective. This question is ultimately quite important to providers and payers, because it addresses whether a new intervention will work at a given community-based treatment program, or for a subset of patients at that program, or whether modifications to usual care at the program are warranted that would synergize with the new intervention to produce better overall outcome.

Handling sites in the design and statistical analysis of multi-site effectiveness trials is a complex topic, covered in detail elsewhere (Brown and Prescott, 1999; Raudenbush & Liu, 2000; Mikulich, Zerbe, & Feaster, 2003; Feaster, Robbins, Horigian, & Szapocznik, 2004). Designs 1, 2a, and 2b may be more likely to observe site variability, because the control condition is simply TAU, and there is no effort to control the treatment delivered. Designs 3 and 4, by specifying the control condition to varying degrees, may reduce site variability. Designs also need to specify how many sites, the sample size per site, how sites will be selected, and whether site is entered as a fixed or random effect in the statistical model estimating the treatment effect. In multi-site designs, getting an accurate estimate of site variability depends, in part, on getting participation from a sample of clinical sites that is representative of the larger population of sites. In efficacy research, one is used to thinking about the representativeness of the sample of patients recruited (and that sample may be deliberately restricted), but for multisite effectiveness trials, the representativeness of the sample of sites is an added dimension.

One approach is to recruit a large number of sites, hopefully a representative sample of sites, and treat site as a random effect in the statistical model used to estimate the treatment effect. The ideal would be random selection of sites, but in reality selection of sites is far from random, since sites in a network like the CTN are highly preselected for their willingness to participate in research, stability of staffing patterns, and perhaps even for their past performance on research projects. Here, the number of patients per site can be relatively modest. As long as the number of sites is large, this will yield an estimate of treatment effect that can, in theory be generalized to the larger population of clinical sites. The site by treatment interaction can also be estimated across all sites, although with small sample sizes per site there would not be power to examine differences between individual sites. Generally, at least 20 sites would be recommended for this approach, and as the data will show (see Table 2), CTN trials fall short of this, generally involving somewhere between 4 to 12 sites.

An alternative approach is to select a smaller number of sites, and recruit a larger number of patients at each site, so that there is sufficient statistical power to approach detecting the treatment effect within each site. Larger per site sample sizes provide more power to detect differences between specific individual sites, either overall difference in outcome (site main effects), or differences in treatment efficacy between sites (site by treatment interactions). Two CTN studies, CTN 0004 (Carroll et al., 2006) and CTN 0005: Ball et al., 2007) planned for 100 participants per site, powered to detect medium-sized intervention effects within each site.

Performance of Effectiveness Designs in the Clinical Trials Network

Table 2 and Table 3 summarize the designs and outcomes, respectively, of each of the CTN effectiveness trials published as of December 2009, grouped according to the type of design represented. Designs 2a, 2b, and 3 are represented, as well as a 3-group study (Campbell et al., 2009), which can be viewed as a hybrid of Designs 2 and 3. None of the studies published to date have used Design 1 or 4. Data were extracted from the primary outcome papers, and in some instances from prior publications describing the study rationales and designs. Data extracted and summarized were intended to bear on the six methodological issues outlined above (see also Table 1), and include, in addition to the design type, the new interventions being tested and their control groups, the clinical setting and patient population, number of sites and sample size per site (Table 2), and the principal outcomes, effect sizes, and effects of site (Table 3). Effect sizes were calculated as Cohen's d (the standardized difference between means) for continuous outcome measures, and Cohen's h (the difference between the arcsine transformations of proportions) for dichotomous outcome measures. Cohen's d and h are roughly comparable, with a value of 0.20 representing a "small" effect, 0.50 a "medium" effect, and 0.80 a "large" effect (Cohen, 1988). For the most part, effect sizes were calculated from the raw means and standard deviations, or the raw proportions. A few of the papers reported effect sizes, and in a few cases effect sizes were calculated from the test statistics. The effect sizes were compiled for descriptive purposes, and formal meta-analytic procedures for combining effect sizes were not attempted, nor did they seem appropriate.

Design 2a: New intervention + TAU versus TAU; New intervention is an add-on to TAU

CTN 0006, and CTN 0007—These two parallel studies tested the effectiveness of a low cost prize incentive behavioral therapy, with prizes contingent on drug negative urines, among stimulant abusers in outpatient drug treatment (CTN 0006), and in methadone maintenance (CTN 0007). The prize incentive intervention was layered on top of treatment as usual (TAU) and administered by research assistants or therapists. Design 2a (New Intervention + TAU vs TAU) appears well justified in regard to the question addressed, since the low cost prize incentive condition is inherently an add-on to TAU, and had strong prior efficacy data (Petry, 2000; Petry et al., 2004; Petry et al., 2005b; Petry & Martin, 2002). The alternative design choice for an add-on treatment, Design 3 (New intervention + TAU vs Control + TAU; with the attention Control being non-contingent prizes or an alternative extra counseling activity), would have parsed out the non-specific elements of the intervention (attention or Hawthorne effects), but this would seem of less interest given the efficacy data that has already clearly shown that incentives have specific effects over and above attention-based controls (Petry et al., 2004; Petry et al., 2005b; Petry & Martin, 2002).

As can be seen in Table 3, the effect sizes on drug use outcome were in the medium range, consistent with the prior efficacy studies. A robust effect is also consistent with Design 2a, where the effect estimated is the sum of the specific and non-specific elements of the treatment. Neither drug use outcome nor attrition in the TAU group seemed so poor as to suggest a Wait-list control effect. Inspection of the data suggests some variability in overall outcome by site, and differences in effect size across sites. However, the effects were in the same direction (prizes + TAU superior to TAU alone) for all sites, suggesting that effects of site and site by treatment, if present, were small. Formal analysis of effects of site, and site by treatment, were not reported, but would have been limited by modest per site sample sizes and numbers of sites.

CTN 0009—This study tested the effectiveness of an intervention for nicotine dependence, nicotine patch plus a cognitive behavioral group intervention, when offered on top of usual

substance abuse treatment (Reid et al., 2008). Two types of treatment programs were included, outpatient "drug-free" programs and methadone maintenance programs, although most recruitment occurred at the latter. Design 2a (New Intervention + TAU vs TAU) continues to appear appropriate in retrospect in regard to the question addressed. Nicotine patch plus counseling has solid efficacy data. Most community-based programs offer little in the way of smoking cessation, so that such treatment would be, in fact, an add-on. The effect sizes observed were in the medium range. However, for the most important outcome, smoking abstinence, the rate was very low in the intervention group (10% during treatment, 6% after treatment completion), and the medium effect size is a function of the negligible abstinence rate (0%) in the control condition. Effects of site (other than difficult recruitment at the outpatient sites) were not reported but would be of limited meaning given the low response rates and modest per site sample sizes and site number. Overall, this study was useful, suggesting that smoking cessation research is feasible among drug treatment patients, and that more powerful interventions need to be tested, such as bupropion or varenicline.

Design 2b: New Intervention + TAU vs TAU; New intervention substitutes for part of TAU

CTN 0004, CTN 0005, and CTN 0021—These studies tested the effectiveness of 1 session of Motivational Interviewing (MI) (CTN 0004: Carroll et al., 2006) or 3-session Motivational Enhancement Therapy (MET) (CTN 0005: Ball et al., 2007) integrated into the early weeks community-based drug treatment. CTN 0021 (Carroll et al., 2009) tested MET among Spanish speaking patients with a design parallel to that of CTN 0005. MI and MET stress a collaborative, non-confrontational approach to motivating patients, which is the antithesis of a more directive, confrontational stance typical when clinicians deal with ambivalent patients (Miller & Rollnick, 2002). The question addressed by Design 2b seems the right one, since the Motivational approach is fundamentally a substitute for usual counseling practices, and there are ample prior efficacy data. As can be seen in Table 3, effects were small, but were significant for some outcomes, including a sleeper effect in CTN 0005, as often observed in studies of psychotherapeutic treatments for substance abuse. The small effect size is consistent with modest effects observed in prior efficacy data, and the fact that the intervention itself is brief. Smaller effect sizes may also be expected with Design 2b, because it controls for time with clinicians. In both CTN 0005 and CTN 0021, effects of MET were larger, and statistically significant among the subgroup of patients with alcohol as the primary substance of abuse. The outcomes in the TAU control conditions are reasonably good (e.g. one month treatment retention in the 70% to 80% range for CTN 0004; 16-week retention in the 40% range for CTN 0005; 12-week retention in the 50% range for CTN 0021; low rates of substance use), reducing concerns about Wait-list effects.

CTN 0004 and CTN 0005 are unique among CTN studies in calling for a smaller number of sites (5 each), and sample sizes of 100 patients per site in an effort to power each site to be able to independently detect effects. Similarly, CTN 0021 targeted 80 patients per site. Correspondingly, CTN 0005 and CTN 0021 detected main effects of site on all outcomes of substance use and retention. In CTN 0005, a site by treatment interaction on one of the outcomes (proportion positive urines) was interpreted with caution, because it was a non-significant trend, and the pattern was not consistent across outcomes. In all three studies, the sites were quite heterogeneous in patient characteristics, and in CTN 0005 sites differed in the number of ancillary sessions delivered as part of TAU and in the duration of study sessions. CTN 0004 detected variations between sites in the skill achieved by the therapists. MI is not easy to learn, and clinicians vary in skill levels achieved after training (Miller & Rollnick, 2002; Miller & Mount, 2001). A substantial accomplishment of these studies is that they trained most participating clinicians to adequate levels of proficiency. Nonetheless, important questions remain for the field in terms of how best to train and maintain MI

skillfulness, and what are the ingredients of MI most related to its effect (Amrhein, Miller, Yahne, Palmer, & Fulcher, 2003).

CTN 0013—This study was similar in design to CTN 0005, except that the patient population was pregnant women enrolled in specialized treatment programs (Winhusen et al., 2008). The version of MET tested was adapted to address the issues unique to pregnant women, such as the future health of their offspring as a motivator. Given that prior efficacy data were limited for this version of MET (Jones, Svikis, & Tran, 2002; Jones, Svikis, Rosado, Tuten, & Kulstad, 2004), Design 3 with a specified control condition might have been chosen. However, the design team, particularly the practitioners at the CTN-affiliated programs for pregnant women, felt that the pragmatic question addressed by Design 2b was most clinically relevant.

Recruitment was challenging, and sample sizes varied by site with smaller per-site sample sizes than for CTN 0004 and CTN 0005 (see Table 2). Nonetheless, the overall sample size of 200 represents the largest randomized, controlled trial in pregnant substance users to date and shows the ability of the CTN to conduct research with important populations that are difficult to recruit in large numbers at any one site. As can be seen in Table 3, the study detected no overall effects of MET compared to TAU. Participants in both conditions reported significant decreases in alcohol/drug use during the first month of treatment, and this may be an example of the effect of a brief intervention (MET) being overwhelmed by an intensive and relatively potent TAU in the specialized programs for pregnant women (Winhusen et al., 2008). There were significant site by treatment interactions, but no clear differences between MET and TAU at any one site. This study provides another example of the potential importance of powering effectiveness trials to detect differences between individual sites, but also the difficulty achieving large per-site sample sizes that would be needed to do so, especially when recruiting samples selected for special clinical attributes (e.g., pregnancy, or comorbid conditions).

Design 3: New Intervention + TAU versus Control Intervention + TAU

CTN 0001, and CTN 0002—These parallel studies tested the effectiveness of the high affinity opiate receptor partial agonist buprenorphine for detoxification from opiates, compared to the non-narcotic clonidine, in community-based residential treatment settings (CTN 0001) and outpatient treatment programs (CTN 0002) (Ling et al., 2009). Design 3 seems appropriate because a detoxification agent such as buprenorphine is best used as a component of some more comprehensive treatment. The clonidine control condition systematizes usual treatment of non-narcotic detoxification. Allowing non-narcotic detoxification methods to vary according to local practices, as in Design 2, would have likely introduced substantial variation.

As can be seen in Table 3, the effect size on the primary outcomes of retention-abstinence and opioid withdrawal symptoms were large, more so in the residential settings. Thus, this study provided a convincing demonstration of effectiveness and helped to introduce this promising new agent to the field. The poor treatment retention and abstinence rates in the clonidine conditions suggest a possible wait list control effect. Clonidine by itself is of some effectiveness for reducing opioid withdrawal symptoms, but the knowledge that a powerful agonist treatment was available in the alternative condition might have caused patients to flee treatment if assigned to clonidine.

The sample sizes of individual sites were too small to readily detect differences between individual sites. However, the higher abstinence rates and the greater effect size for buprenorphine in CTN 0001, compared to CTN 0002, suggests an interaction with type of treatment program, such that buprenorphine detoxification is more effective when used in a

residential setting. The relapse rate after detoxification from opioids is known to be high, and the structure afforded by residential settings may oppose this effect.

CTN 0003—This study addressed the common belief among clinicians that a gradual rate of medication taper is more effective as a detoxification strategy than a rapid taper. After one month of stabilization on buprenorphine in community-based outpatient treatment programs, opioid dependent patients were randomly assigned to either a 25-day taper, or a 7-day taper (Ling et al., 2009). This is a variation on Design 3 in which two forms of a new intervention strategy are compared to determine which is optimal. Internal validity is strong since both interventions are new and plausible and likely surrounded by positive expectations. External validity is also strong since the interventions were delivered within community-based treatment programs.

Contrary to expectations, there was a small advantage for the brief taper in terms of abstinence at the 1-month follow-up point. There was no difference in withdrawal symptoms, which were low in both conditions. However, what is most striking is the very low rate of opioid abstinence by 3 months after completing the taper in both conditions. As in CTN 0002, this result shows the poor overall effectiveness of opioid detoxification when followed by outpatient drug-free treatment. Site sample sizes were modest (mean of 47 patients per site), and effects of site were not reported.

CTN 0010—The study examined the effectiveness of stabilization on buprenorphine for treatment of opioid dependence among adolescents and young adults, a serious and growing public health concern (Woody et al., 2008). A survey of adolescent treatment programs, conducted during the design of CTN 0010, suggested that detoxification, often using non-narcotic medications, was the usual treatment approach, with poor outcome. Thus, the decision was made to use Design 3, and to implement a standard 14-day buprenorphine taper as the control condition in addition to community-based psychosocial treatment. This control would be at least minimally effective and likely to engender positive expectations, protecting internal validity.

As can be seen in Table 3, this study produced the largest effect sizes of any CTN study, and is a landmark in the field, showing for the first time the effectiveness of a buprenorphine stabilization strategy among adolescents and young adults, averaging 19.1 years of age. At the same time, it is notable that the abstinence rates in the buprenorphine-taper control condition at 4, 8, and 12 weeks are higher than the abstinence rate at 3 month follow-up among the adults in CTN 0003. This relatively good outcome in the control condition makes a wait-list control effect seem unlikely. Clinically, it suggests buprenorphine taper followed by drug-free treatment may be an effective approach for a subgroup of opioid dependent adolescents, and future research should seek to identify the clinical characteristics that predict whether adolescents need buprenorphine stabilization, or could succeed with a detoxification and drug free treatment. Effects of site, and site by treatment, which could begin to address these questions, were tested and not detected. The small sample sizes at each site (median N of 30) would limit power to detect difference in effect between individual sites.

CTN 0011—This study examined the effectiveness of telephone follow-ups after discharge from residential treatment as a way to improve adherence with recommended referrals to outpatient treatment programs in the community, and to improve drug use outcome (Hubbard et al., 2007). Consistent with Design 3, all study participants received a standardized discharge planning session prior to discharge from residential treatment then were randomized to either receive follow-up telephone calls over a period of months after discharge, or the control condition that received only the discharge planning session. This

control condition serves to standardize usual care in regard to discharge planning, and the design isolates the effect of the phone calls, strengthening internal validity.

As can be seen in Table 3, the observed effect was weak with at best a small effect favoring telephone follow-up on attendance as documented through records at the outpatient programs. Overall outcome was at least fair, with 45% of patients in the control condition with documented attendance at outpatient treatment and 67% self-reporting attendance. This, and the weak separation at best between intervention and control conditions suggests neither positive effects of attention, nor Wait-list effects. The discrepancy between self-report and documented attendance is a reminder of the importance of attention to measurement issues and of obtaining objective confirmation of outcome when possible. Site by treatment analyses were not reported, although with per site sample sizes of 60 at two sites and 120 at two sites, such analyses would be feasible. Altogether, the findings are useful in suggesting that the telephone intervention tested was by itself too weak to have a useful effect, and more powerful interventions for securing follow-up need to be developed.

CTN 0015—The study tested the effectiveness of Seeking Safety, a cognitive-behavioral treatment designed for women with PTSD and substance dependence, compared to an attention control condition, Women's Health Education (Hien et al., 2009). Seeking Safety could have been implemented as a stand-alone treatment, and tested with Design 1 or Design 4. However, treatment focused on a co-occurring psychiatric disorder seemed most appropriate as an add-on to a larger drug treatment effort. The choice of Design 3, with a strict attention control condition, was driven by the fact that prior efficacy studies for Seeking Safety were limited, and that Seeking Safety had been shortened and modified from a longer, individual psychotherapy to a 12-session group therapy. Thus, the design selected had strong internal validity to assess whether the specific elements of Seeking Safety impacted treatment outcomes.

As summarized in Table 3, there were no significant differences between conditions in the primary outcome analyses. However, both conditions showed substantial and clinically significant improvements in PTSD symptoms from baseline. Both conditions were well received by clinicians and patients, generating both strong treatment alliance scores, and esprit de corps among the therapists. In retrospect, positive expectations surrounding both the intervention and control conditions may have generated a Hawthorne effect that overshadowed specific effects of Seeking Safety. Further, the Women's Health Education control, with its emphasis on bodily functioning and health, may have had its own unique efficacy among traumatized, drug dependent women. This is a situation where, in retrospect, it would have been advantageous to add a TAU control condition in a 3-group hybrid combining Design 2 and Design 3. The TAU control would have addressed the impact of adding more treatment of either intervention type, compared to current usual practice. Practitioners wanted to know what to do to improve outcome in these patients, and the 2-group design with strict attention control failed to answer this practical question. A subsequent secondary analysis did show that improvement in PTSD was associated with subsequent improvement in substance abuse (Hien et al., 2010), more so among those who received Seeking Safety and had more severe substance use at baseline. Effects of site were detected, but no site by treatment interactions.

CTN 0018, and CTN 0019—These studies tested gender-specific, skills oriented interventions to reduce HIV risk behavior in men (CTN 0018; Calsyn et al., 2009), and women (CTN 0019; Tross et al., 2008) enrolled in community-based substance abuse treatment (8 methadone and 8 psychosocial treatment). The interventions were delivered over five 90-minute group sessions. The intervention for women, Safer Sex Skills Building (SSB), was supported by a prior efficacy trial (El-Bassel & Schilling, 1992). The

intervention for men, Real Men are Safe (REMAS), was modeled in part after Project Light (National Institute of Mental Health Multisite HIV Prevention Trial Group, 1998) and Time Out for Men (Bartholomew & Simpson, 1996), but, as a package, had less prior efficacy data.

Design 3 was chosen. The control condition was a single 60-minute HIV Education session (plus TAU), manual-guided, consisting of a subset of the material provided in the experimental conditions. This was intended to standardize what is usually offered in community-based clinics (Shoptaw et al., 2002). Standard HIV education practices varied between community-based treatment programs and could be quite minimal. The control session was intended to reduce variability between programs, and to provide a credible basic intervention, improving internal validity in that both intervention and control conditions were of high quality and would be delivered with enthusiasm by the same interventionists, thus eliminating possible therapist effects. The conditions were not balanced on quantity of time, as this was inherently in part a test of "more is better".

For both studies, the interventions significantly reduced HIV risk behavior, compared to control, across 3 and 6 month follow-up points (Table 3). The effect is medium sized for the women's intervention (CTN 0019), and smaller for the men's intervention (CTN 0018). The studies had a relatively large number of sites (14 for CTN 0018, and 12 for CTN 0019) (see Table 2) and modest numbers of participants per site, averaging in the low 40s. Site was entered as a random effect into the statistical models testing efficacy.

Design 2/3 Hybrid: New Intervention + TAU versus Alternate Intervention + TAU versus TAU

CTN 0017—The first aim of this study was to test the effectiveness of a two session counseling and education intervention aimed at reducing risk for transmission of HIV and hepatitis C, among injection drug users. A second aim was to test a brief Therapeutic Alliance intervention to address the problem of failure to enroll in outpatient treatment after an index episode of inpatient detoxification for injection drug users. The data on treatment adherence after discharge are published (Campbell et al., 2009). In this 3-arm study, usual treatment (TAU) at community-based residential detoxification units was the control condition. The presence of two new interventions, roughly comparable on quantity and quality of attention but with different aims, can be viewed as creating a hybrid design combining Design 2a (Therapeutic Alliance intervention + TAU vs TAU), and Design 3 (Therapeutic Alliance + TAU vs HIV/HCV counseling and education). In Table 2 and Table 3, CTN 0017 is broken down accordingly into its Design 2a (first row in the Tables under CTN 0017) and Design 3 (second row under CTN 0017) components. Viewed this way, the design provided both a more rigorous test of efficacy of the specific elements of the Therapeutic Alliance intervention with strong internal validity (compared to the HIV/HCV intervention), and also addresses the more practical question of its net effectiveness over usual treatment.

The effect sizes observed are in the small range, not surprising for brief interventions. The effect of site was significant for the outcome of outpatient treatment entry, suggesting variability across sites overall, but no site by treatment effect was reported. What is of particular methodological interest is that in the more rigorous test of efficacy (Table 3, CTN 0017, second row: TA intervention vs HIC/HCV intervention, consistent with Design 3), the effect of TA on the primary outcome measure (probability of outpatient treatment entry) is smaller (0.13) and falls short of significance. Whereas, when TA is contrasted to TAU (Table 3, CTN 0017, first row, consistent with Design 2a) the effect size on outpatient treatment entry is larger (0.22) and statistically significant. This is the same problem with Design 3 encountered by CTN 0015, namely without a TAU control condition the overall

effect on the outcome compared to TAU is not measured, and this may be the question of most interest to providers and payors. This suggests the value of the Design 2/3 hybrid illustrated by CTN 0017, and suggests such 3-arm designs should be considered more seriously for community-based effectiveness trials.

Discussion

First, this review serves to illustrate how much has been accomplished by the National Drug Abuse Clinical Trials Network during its first 10 years. Across the first 16 trials for which the primary outcome analyses are published, 5,958 patients with substance use disorders were randomized, across 114 community-based treatment programs located across the United States. The trials tested a range of new evidence based interventions, including buprenorphine and various behavioral interventions aimed at substance use, HIV risk, and an important co-occurring psychiatric disorder, PTSD, and provided findings with important implications for the delivery of community-based treatment for substance use disorders. The main purpose of this review paper was to analyze the experimental designs chosen for each of the published CTN trials and how well they served the goals of effectiveness research. We submit that several main conclusions, or lessons learned may be drawn from this analysis.

Addressing the right questions

Effectiveness research seeks primarily to address to what extent better outcome results from a new intervention, compared to the current standard of practice in the community. The CTN studies reviewed examined new interventions that are add-ons to TAU, or substitute for portions of TAU. New interventions were compared either to treatment as usual (TAU) unmodified (Designs 2a and 2b) or to some specified control condition integrated into TAU and intended to systematize and provide a credible model for usual practice, and to reduce noise from site variability (Design 3). The impression on balance is that these designs addressed the right questions and were thus useful to the field as intended. Only one CTN study so far uses a design where the new intervention is a complete substitute for standard treatment (Designs 1) (Robbins et al., 2009), and the primary outcomes are not yet published. Such comprehensive new interventions may be relatively unusual, given the breadth of current intervention programming in community-based treatment programs and the wide-ranging service needs of their clinical populations.

Internal versus external validity

Conducting the CTN studies at community-based treatment programs, recruiting representative patients from those programs, and having community-based practitioners trained and conducting the interventions are all in the service of external validity. By having the control conditions reflect TAU, the designs themselves address the question most important to the field of community-based treatment, as noted above. We considered threats to the internal validity of these studies including failure to control for the quantity or quality of attention and expectancy effects provided by the new interventions being tested.

So called Wait-list control effects may be the most important threat, because they may create poor outcome in the control condition that is an artifact of study participation. While a more complicated experimental design would be needed to rigorously assess for this, Wait-list effect would be suggested by particularly poor outcome or high drop out in the control condition. Fortunately, there was relatively little evidence for this among the CTN trials reviewed. The buprenorphine detoxification studies, where clonidine was the control condition (CTN 0001 and CTN 0002), observed poor outcome on clonidine, although this is not a great surprise given other existing evidence on the limited effectiveness of non-

narcotic detoxification agents, and the poor long term effectiveness of opioid detoxification itself.

It may be that study participation itself creates a positive atmosphere that opposes wait-list effects and provides some base of positive expectations for all participants. At a minimum study participation involves regular visits with enthusiastic research staff for measurement and likely some modest incentives for participation. There is always the concern that measurement effects could overwhelm intervention effects if the time spent with research staff for assessment is extensive and exceeds time in treatment. This suggests measurement for effectiveness trials should be lean, but should create a positive atmosphere around study participation. It may also be useful to measure patients' expectations about study treatments at baseline.

Effect size and power

More than half of the CTN studies detected significant intervention effects. Looking across studies in Table 3, the impression is that magnitude of effects observed was related mainly to the potency of the interventions. Buprenorphine, a powerful medication treatment for opioid dependence (CTN 0001, CTN 0002, CTN 0010), and incentives, arguably the most powerful behavioral intervention in the current armamentarium (CTN 0006, CTN 0007), generated the largest effect sizes. Smaller effects detected in other studies were consistent with estimates from the prior efficacy studies. This is encouraging in suggesting that Stage 2 efficacy studies are useful as a guide both to selection of potentially effective interventions and estimation of sample size. This provides some validation for the Stage model of treatment development that has been the basis of the treatment development efforts of NIDA (Rounsaville, Carroll, & Onken, 2001). That said, most interventions in behavioral health in general have effects that are small, or in the range between small and medium, and for such effects, even sample sizes in the ranges utilized by the CTN studies (N = 300 to 500) may be only marginally adequate. Design strategies to enhance power, and larger sample sizes, should be considered in future studies. Ultimately, judgment from clinical, public health, and cost perspectives should guide determination of whether small effect sizes are meaningful, likely to be used by treatment providers, and worth the cost.

More consideration for 3-arm designs

Three arm designs have been largely shunned by the CTN because of the greater logistical complexity of implementing three experimental conditions at community-based sites, and the added sample size and expense. However, the studies reviewed include at least one example, CTN 0015, where in retrospect a three arm design would have been preferable. CTN 0015 implemented a rigorous attention control condition for the new intervention being tested, Seeking Safety, because of concerns about limited efficacy data and the fact that Seeking Safety had been significantly modified into a shorter group intervention to make it more feasible for community-based treatment settings. The control condition, Women's Health Education, was associated with large reductions in PTSD symptoms similar to those observed with Seeking Safety, and the principal outcome analysis was negative. Addition of a third arm, with TAU alone, or TAU plus some minimal control intervention would have answered the question of whether both of the other interventions improved outcome compared to standard practice. CTN 0017 provided an example of a 3-arm design with both attention control and TAU control conditions, and, as expected it was the contrast with the TAU control where larger estimates of effect size were observed.

Three arm designs might also be considered to assess the relative impact of two different levels of an intervention, compared to TAU. When the observed effect size is small, it begs the question as to how the intervention could be modified to produce a larger effect, either

by providing more of the intervention, or adding some new component. An example might be combinations of prize incentives with psychotherapeutic interventions such as MI or cognitive behavioral and skills oriented interventions.

More attention to the effects of site

This is a theme that emerged from our examination of many of the CTN trials. Not all the trials reported analysis of the effects of site, but among those that did, most detected either main effects of site, or site by treatment interactions either on outcome measures, or on measured skill of the clinicians delivering the study interventions. This is not surprising, since community-based treatment programs vary widely in their patient populations, clinicians, and usual practices (TAU). Site by treatment interactions help explain the conditions under which a new intervention is more or less successful. Differential effectiveness across sites begins to address whether the behavioral interventions that are the backbone of most community-based programs are more effective at some programs than others, and this could provide hypotheses about how community-based treatment could be improved. It may also reflect differences in patients populations or resources available in the community. These seem very important questions for effectiveness research to address, and they have generated considerable interest amongst the community-based clinicians and providers participating in the CTN.

As noted previously, there are two broad approaches to sites in the design of multisite effectiveness trials. One is to recruit a large number of sites, preferably 20 or more; here per site sample size can be relatively small; overall effects of site and site by treatment can be estimated, but differences among particular sites are harder to detect. Another is to select a smaller number of sites, with each site adequately powered to detect an effect (generally 100 per site or more depending on the expected effect size); this provides more ability to relate site effects to specific site characteristics. Inspection of Table 2 shows that most studies fell in between these two approaches, with per site sample sizes less than 100 and number of sites of 10 to 12 or less. Choice of site number and per site sample size is often driven by logistical concerns, namely given the overall target N, how many sites would be needed to finish the study recruitment quickly. To achieve larger per site sample sizes, studies might need to keep recruitment open for longer; the disadvantage would be slower study completion and slower dissemination of results to the field. Larger per site sample sizes would also bias against inclusion of smaller community-based treatment programs, reducing the representativeness of sites. Studies may need to be larger overall if individual sites are to be adequately powered, and a sufficiently representative sample of sites included. Larger studies will cost more, and efforts are needed to increase efficiency and make the most out of scarce research dollars.

Concerning a related issue, as the CTN matures, there may be a tendency for sites to be selected for participation in new trials based upon their track record in performing prior research studies, rather than for their representativeness of the larger population of community-based treatment programs. This emphasis is understandable, as it protects the integrity and feasibility of the research. But, over time, these select community-based treatment programs may become more like research clinics, undermining the precept that effectiveness research be conducted in real world settings.

More attention to cost-effectiveness

When is a given effect worth detecting, and when is a treatment with a given effect size worth implementing in the field? These questions can be addressed by surveys of providers (Miller & Manuel, 2008). However, analyses of cost-effectiveness and cost benefit in large randomized effectiveness provide uniquely powerful data. Only a few of the CTN studies

published to date, including CTN 0006 (Olmstead, Sindelar, & Petry, 2007), CTN 0007 (Sindelar, Olmstead, & Peirce, 2007), and CTN 0010 (Polsky et al., submitted—Personal Communication, G. Woody), have reported such analyses or included them among their primary aims. The question of what size effect is worth detecting is directly relevant to design in terms of whether a large effectiveness trial is worth doing, and if so the choice of sample size. Program directors and payors may be impressed by evidence of clinical effectiveness, but still unsure of whether the effort to implement the new intervention is worth the cost. Also, increased costs of delivering a new intervention in the short run may be offset by long-term reductions in costs related to improved long term outcome, such as social, medical, or criminal justice costs.

Finally, it is important to note that this review has restricted itself to examining basic experimental designs for effectiveness trials in terms of choice of control groups and sites, and that other important methodological issues have not been addressed. Among others, these would include whether to randomize clinicians to intervention conditions, or randomize patients within clinicians, or when to consider randomizing entire programs to different regimens, and issues surrounding methods of training and supervising clinicians in new interventions. We have not considered more complex adaptive designs, nor designs that examine methods of training and dissemination that bear on the likelihood that a new treatment will be adopted in an effective manner. Nonetheless, it is hoped that the paper has shown how much can be accomplished through a concerted effort to fund research on treatment effectiveness, and the rich potential for future advances informed by the lessons learned from this first generation of studies in the Clinical Trials Network.

Acknowledgments

References

Amrhein PC, Miller WR, Yahne CE, Palmer M, Fulcher L. Client commitment language during motivational interviewing predicts drug use outcomes. Journal of Consulting and Clinical Psychology 2003;71:862–878. [PubMed: 14516235]

Ball SA, Martino S, Nich C, Frankforter TL, Van Horn D, Crits-Christoph P, Woody GE, Obert JL, Farentinos C, Carroll KM. National Institute on Drug Abuse Clinical Trials Network. Site matters: multisite randomized trial of motivational enhancement therapy in community drug abuse clinics. Journal of Consulting and Clinical Psychology 2007;75:556–567. [PubMed: 17663610]

Bartholomew, NG.; Simpson, DD. Time Out! For Men: A Communication Skills & Sexuality Workshop for Men. Fort Worth: Texas Christian University, Institute of Behavioral Research; 1996.

Brigham GS, Feaster DJ, Wakim PG, Dempsey CL. Choosing a control group in effectiveness trials of behavioral drug abuse treatments. Journal of Substance Abuse Treatment 2009;37:388–397. [PubMed: 19553062]

Brown, H.; Prescott, R. Applied Mixed Models in Medicine. London: John Wiley & Sons; 1999.

Calsyn DA, Hatch-Maillette M, Tross S, Doyle SR, Crits-Christoph P, Song YS, Harrer JM, Lalos G, Berns SB. Motivational and skills training HIV/sexually transmitted infection sexual risk reduction groups for men. Journal of Substance Abuse Treatment 2009;37:138–150. [PubMed: 19150206]

Campbell BK, Fuller BE, Lee ES, Tillotson C, Woelfel T, Jenkins L, Robinson J, Booth RE, McCarty D. Facilitating outpatient treatment entry following detoxification for injection drug use: a multisite test of three interventions. Psychology of Addictive Behaviors 2009;23:260–270. [PubMed: 19586142]

Carroll KM, Ball SA, Nich C, Martino S, Frankforter TL, Farentinos C, Kunkel LE, Mikulich-Gilbertson SK, Morgenstern J, Obert JL, Polcin D, Snead N, Woody GE. National Institute on Drug Abuse Clinical Trials Network. Motivational interviewing to improve treatment engagement and outcome in individuals seeking treatment for substance abuse: a multisite effectiveness study. Drug and Alcohol Dependence 2006;81:301–312. [PubMed: 16169159]

Carroll KM, Martino S, Ball SA, Nich C, Frankforter T, Anez LM, Paris M, Suarez-Morales L, Szapocznik J, Miller WR, Rosa C, Matthews J, Farentinos C. A multisite randomized effectiveness trial of motivational enhancement therapy for Spanish-speaking substance users. Journal of Consulting and Clinical Psychology 2009;77:993–9. [PubMed: 19803579]

Cohen, J. Statistical power analysis for the behavioral sciences. 2nd edition. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers; 1988.

El-Bassel N, Schilling RF. 15-month followup of women methadone patients taught skills to reduce heterosexual HIV transmission. Public Health Reports 1992;107:500–504. [PubMed: 1410230]

Feaster DJ, Robbins MS, Horigian V, Szapocznik J. Statistical issues in multisite effectiveness trials: the case of brief strategic family therapy for adolescent drug abuse treatment. Clinical Trials 2004;1:428–439. [PubMed: 16279281]

Hien DA, Wells EA, Jiang H, Suarez-Morales L, Campbell AN, Cohen LR, Miele GM, Killeen T, Brigham GS, Zhang Y, Hansen C, Hodgkins C, Hatch-Maillette M, Brown C, Kulaga A, Kristman-Valente A, Chu M, Sage R, Robinson JA, Liu D, Nunes EV. Multisite randomized trial of behavioral interventions for women with co-occurring PTSD and substance use disorders. Journal of Consulting and Clinical Psychology 2009;77:607–619. [PubMed: 19634955]

Hien DA, Jiang H, Campbell ANC, Hu M, Miele GM, Cohen LR, Brigham GS, Capstick C, Kulaga A, Robinson J, Suarez-Morales L, Nunes EV. Do treatment improvements in PTSD severity affect substance use outcomes? A secondary analysis from a randomized clinical trial in NIDA's Clinical Trials Network. American Journal of Psychiatry 2010;167:95–101. [PubMed: 19917596]

Hubbard RL, Leimberger JD, Haynes L, Patkar AA, Holter J, Liepman MR, Lucas K, Tyson B, Day T, Thorpe EA, Faulkner B, Hasson A. National Institute on Drug Abuse. Telephone enhancement of long-term engagement (TELE) in continuing care for substance abuse treatment: a NIDA clinical trials network (CTN) study. American Journal on Addictions 2007;16:495–502. [PubMed: 18058417]

Institute of Medicine. Bridging the gap between practice and research: forging partnerships with community-based drug and alcohol treatment. Washington, D.C: National Academy Press; 1998.

Jones HE, Svikis D, Rosado J, Tuten M, Kulstad JL. What if they do not want treatment?: lessons learned from intervention studies of non-treatment-seeking, drug-using pregnant women. American Journal on Addictions 2004;13:342–357. [PubMed: 15370933]

Jones HE, Svikis DS, Tran G. Patient compliance and maternal/infant outcomes in pregnant drug-using women. Substance Use and Misuse 2002;37:1411–1422. [PubMed: 12371578]

Ling W, Amass L, Shoptaw S, Annon JJ, Hillhouse M, Babcock D, Brigham G, Harrer J, Reid M, Muir J, Buchan B, Orr D, Woody G, Krejci J, Ziedonis D. Buprenorphine Study Protocol Group. A multi-center randomized trial of buprenorphine-naloxone versus clonidine for opioid detoxification: findings from the National Institute on Drug Abuse Clinical Trials Network. Addiction 2005;100:1090–1100. [PubMed: 16042639]

Ling W, Hillhouse M, Domier C, Doraimani G, Hunter J, Thomas C, Jenkins J, Hasson A, Annon J, Saxon A, Selzer J, Boverman J, Bilangi R. Buprenorphine tapering schedule and illicit opioid use. Addiction 2009;104:256–265. [PubMed: 19149822]

Mikulich SK, Zerbe GO, Feaster DJ. Some ramifications of treating "site" as random in multi-center clinical trials. Controlled Clinical Trials 2003;24:43S–2405.

Miller, WM.; Rollnick, S. 2nd Ed. New York: Guilford Press; 2002. Motivational interviewing: Preparing people for change.

Miller WR, Mount KA. A small study of training in motivational interviewing. Does one workshop change clinician and patient behavior? Behavioural and Cognitive Psychotherapy 2001;29:457–471.

Miller WR, Manuel JK. How large must a treatment effect be before it matters to practitioners? An estimation method and demonstration. Drug and Alcohol Review 2008;27:524–528. [PubMed: 18608445]

National Institute of Mental Health (NIMH) Multisite HIV Prevention Trial Group. The NIMH Multisite HIV Prevention Trial: reducing HIV sexual risk behavior. Science 1998;280:1889–1894. [PubMed: 9632382]

Olmstead TA, Sindela JL, Petry NM. Clinic variation in the cost-effectiveness of contingency management. Am J Addict 2007;16:457–60. [PubMed: 18058410]

Peirce JM, Petry NM, Stitzer ML, Blaine J, Kellogg S, Satterfield F, Schwartz M, Krasnansky J, Pencer E, Silva-Vazquez L, Kirby KC, Royer-Malvestuto C, Roll JM, Cohen A, Copersino ML, Kolodner K, Li R. Effects of lower-cost incentives on stimulant abstinence in methadone maintenance treatment: a National Drug Abuse Treatment Clinical Trials Network study. Archives of General Psychiatry 2006;63:201–208. [PubMed: 16461864]

Petry NM. A comprehensive guide to the application of contingency management procedures in clinical settings. Drug and Alcohol Dependence 2000;58:9–25. [PubMed: 10669051]

Petry NM, Martin B. Low-cost contingency management for treating cocaine- and opioid-abusing methadone patients. Journal of Consulting and Clinical Psychology 2002;70:398–405. [PubMed: 11952198]

Petry NM, Tedford J, Autsin M, Nich C, Carroll KM, Rounsaville BJ. Prize reinforcement contingency management for treatment of cocaine abusers: How low can we go, and with whom? Addiction 2004;99:349–360. [PubMed: 14982548]

Petry NM, Peirce JM, Stitzer ML, Blaine J, Roll JM, Cohen A, Obert J, Killeen T, Saladin ME, Cowell M, Kirby KC, Sterling R, Royer-Malvestuto C, Hamilton J, Booth RE, Macdonald M, Liebert M, Rader L, Burns R, DiMaria J, Copersino M, Stabile PQ, Kolodner K, Li R. Effect of prize-based incentives on outcomes in stimulant abusers in outpatient psychosocial treatment programs: a national drug abuse treatment clinical trials network study. Archives of General Psychiatry 2005a; 62:1148–1156. [PubMed: 16203960]

Petry NM, Alessi SM, Marx J, Austin M, Tardif M. Vouchers versus prizes: Contingency management for treatment of substance abusers in community settings. Journal of Consulting and Clinical Psychology 2005b;73:1005–1014. [PubMed: 16392974]

Raudenbush SW, Liu X. Statistical power and optimal design for multisite randomized trials. Psychological Methods 2000;5:199–213. [PubMed: 10937329]

Reid MS, Fallon B, Sonne S, Flammino F, Nunes EV, Jiang H, Kourniotis E, Lima J, Brady R, Burgess C, Arfken C, Pihlgren E, Giordano L, Starosta A, Robinson J, Rotrosen J. Smoking cessation treatment in community-based substance abuse rehabilitation programs. Journal of Substance Abuse Treatment 2008;35:68–77. [PubMed: 17951021]

Robbins MS, Szapocznik J, Horigian VE, Feaster DJ, Puccinelli M, Jacobs P, Burlew K, Werstlein R, Bachrach K, Brigham G. Brief Strategic Family Therapy for Adolescent Drug Abusers: A Multi-Site Effectiveness Study. Contemporary Clinical Trials 2009;30:269–278. [PubMed: 19470315]

Rounsaville BJ, Carroll KM, Onken LS. NIDA's stage model of behavioral therapies research: getting started and moving on from Stage I. Clinical Psychology: Science and Practice 2001;8:133–142.

Shoptaw S, Tross S, Stephens M, Tai B. NIDA CTN HIV/AIDS Workgroup. A snapshot of HIV/AIDS-related services in the clinical treatment providers for NIDA's clinical trials network. Drug and Alcohol Dependence 2002;66:S163.

Sindelar JL, Olmstead TA, Peirce JM. Cost-effectiveness of prize-based contingency management in methadone maintenance treatment programs. Addiction 2007;102:1463–71. [PubMed: 17645428]

Tross S, Campbell AN, Cohen LR, Calsyn D, Pavlicova M, Miele GM, Hu MC, Haynes L, Nugent N, Gan W, Hatch-Maillette M, Mandler R, McLaughlin P, El-Bassel N, Crits-Christoph P, Nunes EV. Effectiveness of HIV/STD sexual risk reduction groups for women in substance abuse treatment programs: results of NIDA Clinical Trials Network Trial. Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology 2008;48:581–589.

Winhusen T, Kropp F, Babcock D, Hague D, Erickson SJ, Renz C, Rau L, Lewis D, Leimberger J, Somoza E. Motivational enhancement therapy to improve treatment utilization and outcome in

pregnant substance users. Journal of Substance Abuse Treatment 2008;35:161–173. [PubMed: 18083322]

Woody GE, Poole SA, Subramaniam G, Dugosh K, Bogenschutz M, Abbott P, Patkar A, Publicker M, McCain K, Potter JS, Forman R, Vetter V, McNicholas L, Blaine J, Lynch KG, Fudala P. Extended vs short-term buprenorphine-naloxone for treatment of opioid-addicted youth: a randomized trial. Journal of the American Medical Association 2008;300:2003–2011. [PubMed: 18984887]

Table 1

Summary of experimental designs to test effectiveness of behavioral and pharmacological interventions in community-based treatment settings.

Design 1: New intervention versus Treatment as Usual (TAU)
Primary question addressed:
--How does the effectiveness of a new treatment intervention compare to existing community-based treatment (TAU)?
Suitable interventions to test:
--Treatment approaches that are comprehensive, stand-alone interventions that would substitute entirely for TAU.
Design 2a: New intervention + TAU versus TAU; New intervention is an add-on to TAU
Primary question addressed:
--What is the impact of adding a new treatment intervention to an existing treatment program or TAU?
Suitable interventions to test:
--The intervention is inherently an add-on to an existing treatment program (TAU), not a replacement for part of TAU.
Design 2b: New Intervention + TAU vs TAU; New intervention substitutes for a part of TAU
Primary question addressed:
--What is the impact of substituting a new treatment intervention for an existing component of TAU?
Suitable interventions to test:
--The intervention is inherently a substitute for a component or service currently provided as part of TAU.
Design 3: New Intervention + TAU versus Control Intervention + TAU
Primary question addressed:
--What is the impact of adding a new treatment intervention to existing TAU, over and above the impact of adding a Control intervention? The Control intervention may be designed to control fully for attention, isolating the effect of the specific elements of the new intervention; or, the Control intervention may be designed to standardize some aspect of current TAU.
Suitable treatments to test:
--The treatment should be inherently suitable as an add-on to TAU, or substitute for a component of TAU.
Design 4: New Intervention vs Control Intervention
Primary question addressed:
--What is the impact on outcome of a new treatment intervention compared to a control intervention.
Suitable treatments to test:
--As for Design 1, the treatment is a comprehensive, stand-alone interventions that would substitute entirely for TAU.

Table 2

Summary of design features of randomized clinical trials conducted in the National Institute on Drug Abuse Clinical Trials Network, testing the effectiveness of new treatment interventions among drug dependent patients in community-based drug treatment programs

Study Number Author, Year	New Intervention	Control	Community-based Treatment Setting and Population	Sample Size (N)	Number of Sites	N per Site (range)[a]
Design 2a: [New intervention + TAU] vs TAU; New intervention is an add-on to TAU						
CTN 0006 Petry et al., 2005a	Voucher incentives contingent on abstinence, 12 weeks	TAU	Outpatient, cocaine or other stimulants	415	8	52 (27 – 88)
CTN 0007 Peirce et al., 2006	Voucher incentives contingent on abstinence, 12 weeks	TAU	Methadone maintenance with cocaine or stimulants	388	6	59 (45 – 92)
CTN 0009 Reid et al., 2008	Nicotine patch plus group cognitive behavioral counseling, 12 sessions	TAU	Methadone and outpatient, nicotine dependence	225	7	36
Design 2b: [New Intervention + TAU] vs TAU; New intervention substitutes for a part of TAU						
CTN 0004 Carroll et al., 2006	Motivational Interviewing during baseline evaluation, 1 session	Usual baseline evaluation – 1 session	Outpatient, drugs or alcohol	423	5	100 (23 – 100)
CTN 0005 Ball et al., 2007	Motivational Enhancement Therapy, 3 initial sessions	Usual initial 3 counseling sessions	Outpatient, drugs or alcohol	461	5	100 (61 – 100)
CTN 0013 Winhusen et al., 2008	Motivational Enhancement Therapy for pregnant women, 3 initial sessions	Usual initial 3 counseling sessions	Outpatient, programs for pregnant women, any drugs or alcohol	200	4	58 (10 – 74)
CTN 0021 Carroll et al., 2009	Motivational Enhancement Therapy, 3 initial sessions, delivered in Spanish	Usual initial 3 counseling sessions in Spanish	Spanish speaking outpatients, drugs or alcohol	405	5	81
Design 3: [New Intervention + TAU] vs [Control Intervention + TAU]						
CTN 0001 Ling et al., 2005	Buprenorphine, 14 day taper	Clonidine plus usual counseling	Inpatient or residential, opioid dependent	113	6	19
CTN 0002 Ling et al., 2005	Buprenorphine, 14 day taper	Clonidine plus usual counseling	Outpatient, opioid dependent	231	6	39
CTN 0003 Ling et al., 2009	Buprenorphine, 1 month stabilization then 25 day taper	Buprenorphine 1 month stabilization, then 7 day taper	Outpatient, opioid dependent	516	11	47
CTN 0010 Woody et al., 2008	Buprenorphine 3 month stabilization, then taper, Drug Counseling manual	Buprenorphine 14 day taper, Drug Counseling manual	Adolescents up to 21 years old, opioid dependent	152	6	30 (3 – 52)

Study Number Author, Year	New Intervention	Control	Community-based Treatment Setting and Population	Sample Size (N)	Number of Sites	N per Site (range)[a]
CTN 0011 Hubbard et al., 2007	Discharge planning session, then 4 follow-up phone calls	Discharge planning session	Post residential, drugs or alcohol	339	4	80 (60 – 120)
CTN 0015 Hien et al., 2009	Seeking Safety, 12 group sessions	Women's Health Education, 12 sessions, group	Outpatient, women with PTSD, drugs or alcohol	353	7	50 (7 – 106)
CTN 0018 Calsyn et al., 2009	HIV risk reduction motivation and skills, 5 group sessions	Brief HIV education, 1 group session	Methadone and outpatient, men with risky sex	590	14	42 (11 – 65)
CTN 0019 Tross et al., 2008	Safer Sex Skill Building, 5 group sessions	Brief HIV Education, 1 group session	Methadone and outpatient, women with risky sex	515	12	43
Design 2/3 Hybrid, 3 Group Design: [New Intervention + TAU] vs [Control Intervention + TAU] vs TAU						
CTN 0017 Campbell et al., 2009	Treatment Alliance intervention, 1 session	TAU	Residential detoxification, injection drug users	420	8	53 (26 – 88)
CTN 0017 Campbell et al., 2009	Treatment Alliance intervention, 1 session	HIV/HCV risk reduction, 2 sessions	Residential detoxification, injection drug users	421	8	53 (26 – 88)

[a] N per site is the median, or mean if individual site N's not available

Table 3

Summary of the principal outcomes of randomized clinical trials conducted in the National Institute on Drug Abuse Clinical Trials Network, testing the effectiveness of new treatment interventions among drug dependent patients in community-based drug treatment programs

Study Number Author, Year	Intervention Tested vs Control	Principal Outcome Measures	Intervention Outcome	Control Outcome	Effect Size(a)	Site Effects
Design 2a: [New intervention + TAU] vs TAU; New intervention is an add-on to TAU						
CTN 0006 Petry et al., 2005a	Voucher incentives, 12 weeks vs TAU	Continuous weeks abstinent:	8.6	5.2	0.42***	
		≥ 4 cont. abstinent weeks: In	40%	21%	0.42***	
		treatment at 12 weeks:	49%	35%	0.29**	
CTN 0007 Peirce et al., 2006	Voucher incentives, 12 weeks vs TAU	Continuous weeks abstinent:	5.5	2.3	0.55***	? site, site by treatment
		≥ 4 cont. abstinent weeks:	24%	9%	0.42***	
		In treatment at 12 weeks:	67%	65%	0.04	
CTN 0009 Reid et al., 2008	Nicotine patch + group, 12 sessions vs TAU	Nicotine abstinent week 13:	6%	0%	0.47+	
		Cigarettes per day week 13:	8.8	14.9	0.68***	
Design 2b: [New Intervention + TAU] vs TAU; New intervention substitutes for a part of TAU						
CTN 0004 Carroll et al., 2006	MI during baseline, 1 session vs TAU	Sessions attended month 1:	5.0	4.0	0.24*	Site, site by treatment on skill, ?outcome
		Enrolled at clinic month 1:	84%	75%	0.23*	
		Days using primary substance:	3.3	4.0	0.10	
CTN 0005 Ball et al., 2007	MET, 3 initial sessions vs TAU	In treatment at 16 weeks:	41%	46%	-0.10	Site*** on all outcomes
		Days using primary substance:	~0.2(b)	~0.8	Tx by time**	
CTN 0013 Winhusen et al., 2008	MET-Pregnancy, 3 initial sessions vs TAU	% scheduled hours attended:	62	62	0.00	Site by Tx* on days to drop and drug pos
		Days to dropout:	48	54	-0.17	
		Drug positive urine month 1:	25%	28%	0.07	
CTN 0021 Carroll et al., 2009	MET, 3 initial sessions vs TAU in Spanish	In treatment at 12 weeks:	57%	52%	0.10	Site** on all outcomes
		Days enrolled in treatment:	46.1	42.0	0.15+	
		% days abstinent prim. subst:	94.7	92.2	0.16+	

Study Number Author, Year	Intervention Tested vs Control	Principal Outcome Measures	Intervention Outcome	Control Outcome	Effect Size[a]	Site Effects
Design 3: [New Intervention + TAU] vs [Control Intervention + TAU]						
CTN 0001 Ling et al., 2005	Buprenorphine, 14 day taper vs clonidine	In treatment and opioid negative urine on day 14:	77%	22%	1.17***	
		Opioid withdrawal (COWS):	3.8	7.4	1.35***	
CTN 0002 Ling et al., 2005	Buprenorphine, 14 day taper vs clonidine	In treatment and opioid negative urine on day 14:	29% 4.0	5% 5.1	0.67***	
		Opioid withdrawal (COWS):			0.36**	
CTN 0003 Ling et al., 2009	Buprenorphine, 25 day taper vs 7 day taper	Opioid neg urine post taper:	30%	44%	0.29***	
		Opioid neg urine month 3:	13%	12%	0.03	
		Withdrawal (COWS) p taper:	2.5	2.7	0.06	
CTN 0010 Woody et al., 2008	Buprenorphine × 3 months vs 14 day taper	Opioid negative urine week 4:	74%	39%	0.72***	No site by Tx effects
		Opioid negative urine week 8:	77%	46%	0.65***	
		In treatment at week 12:	70%	21%	1.03***	
CTN 0011 Hubbard et al., 2007	Discharge planning + 4 phone follow-ups vs planning only	≥ 1 outpatient visit self-report:	67%	67%	0.00	
		≥ 1 outpatient visit verified: (from program records)	56%	45%	0.22+	
CTN 0015 Hien et al., 2009	Seeking Safety, vs Women's Health Ed, 12 sessions	PTSD symptoms, self-report (PSSR), months 3, 6, 12:	30.0	32.0	0.14	Site*** on both outcomes, no site by Tx
		Abstinence, months 3, 6, 12:	46%	43%	0.06	
CTN 0018 Calsyn et al., 2009	Safer sex skills, 5 sessions, vs 1 session HIV education	90 day count of unprotected sexual occasions at month 6:	16.0	19.0	0.17*(c)	Site entered as random effect
CTN 0019 Tross et al., 2008	Safer Sex Skills, 5 sessions, vs 1 session HIV Education	90 day count of unprotected sexual occasions at month 6:	14.0	24.1	0.42*(c)	Site entered as random effect
Design 2/3 Hybrid, 3 Group Design: [New Intervention + TAU] vs [Control Intervention + TAU] vs TAU						
CTN 0017	Therapeutic Alliance intervention (TA), 1	Probability of outpatient treatment entry:	0.46	0.31	0.22*	Site*** on outpatient Tx entry
	Therapeutic Alliance intervention (TA), 1 session, vs TAU	Probability of outpatient entry:	0.46	0.65	0.22*	Site*** on outpatient Tx entry

Study Number Author, Year	Intervention Tested vs Control	Principal Outcome Measures	Intervention Outcome	Control Outcome	Effect Size[a]	Site Effects
Campbell et al., 2009		treatment entry:				
CTN 0017 Campbell et al., 2009	TA, 1 session, vs HIV/HCV risk reduction, 2 sessions	Probability of 12-step entry: Probability of outpatient treatment entry:	0.74 0.46	0.65 0.37	0.17 0.13	Site*** on outpatient Tx entry
		Probability of 12-step entry:	0.74	0.63	0.21*	

Significance levels:

+ p < .10;

* p < .05;

** p < .01;

*** p < .001

[a] For continuous measures the effect size estimate is Cohen's d, the standardized difference between means; for dichotomous measures, the effect size estimate is Cohen's h (difference between arcsine-transformed proportions); h and d are comparable.

[b] In CTN 0005, for outcome of days using primary substance, interaction found "sleeper effect" with MET superior in weeks 5 to 16

[c] In CTN 0018 and CTN 0019, treatment effects indicated by significant treatment by time interactions (p < .001)