CHAPTER 25

Multisite Trials in Criminal Justice Settings: Trials and Tribulations of Field Experiments

FAYE S. TAXMAN AND ANNE GIURANNA RHODES

Researchers have choices to make regarding the design of any experiment including the type of design (randomized or quasi-experimental design) and the number of sites to include in a study. The latter decision is one that few have paid significant attention to. As previously argued by Weisburd and Taxman (2000), there are advantages to multicenter trials. Multicenter trials are familiar in medical settings where they provide the opportunity to test a new protocol or innovation within various settings. No decision rules exist as to when a single or multisite trial should occur. Single site trials provide a starting point to test out the feasibility of a new innovation, but multisite trials have a clear advantage in testing the innovation under various operating conditions. In some ways, the multisite trial offers the potential to accumulate knowledge more quickly by using each site as a laboratory for assessing the characteristics or factors that affect the question of "value added" or improved outcomes. The multisite trial creates the opportunity to simultaneously test the innovation as well as learn from these sites about the parameters that affect outcomes. The prospect of doing a multisite trial is more complicated than a single site trial because it requires attention to both the design and the management of the study. It is through multisite trials that we can learn about areas where "drift" may occur, whether it is in the innovation, the management of the study, external factors that affect the innovation, or the inability of the environment to align to the innovation.

This paper is designed to provide insight into the complexity of a multisite trial, and how different design, site preparation, and research management affect the integrity of the experiment. We use case studies of two recently completed multisite trials that involve complex innovations, or innovations that involve more than one operating agency. Unlike other documented work on describing the management problems associated with multicenter trials (a multisite trial with one lead investigator but other site investigators, see Friedman et al. 1985; Pocock 1983), this paper highlights the continuing need to consider implementation issues that affect the integrity of the experiment. In these case studies, we use the various types of threats to internal validity as a conceptual framework to illustrate the decisions that

A.R. Piquero and D. Weisburd (eds.), *Handbook of Quantitative Criminology*,
DOI 10.1007/978-0-387-77650-7_25,

researchers must make in a study. The case studies serve to provide a discussion of issues regarding the design and conduct of multisite trials. We conclude by providing lessons learned in conducting or being involved in multisite trials.

THE MULTISITE TRIAL

While randomized control trials are the "gold-standard" of research, they are often plagued by design and implementation issues. Adding diverse sites complicates matters, particularly when the goal is to develop a standardized protocol across sites, and the study sites have a wide range of expertise and resources. These issues can be even more pronounced in criminal justice settings, where design considerations are often impacted by the structure and administration of local, state and federal agencies. While there has been a push for more trials in the criminal justice field (Weisburd 2003), the number of randomized studies (especially those that are multisite) remains low.

An important first step in multicenter trials is the development of the study protocol. The protocol should have the following: (1) the rationale for the study, (2) a description of the intervention, (3) a list of the procedures to implement the study, (4) copy of the instruments and key measures, and (5) human subject procedures including consent forms and certificates of confidentiality. This protocol is needed to provide the guidance in terms of implementation of the study in the real-world settings with the myriad of organizations, both research organizations and agencies where the innovation is going to be fielded (Chow and Liu 2004). The protocol manual should also consider the unique characteristics of the sites involved, particularly if these characteristics are likely to impact findings or can influence the degree to which the innovation is likely to be delivered. Up front analyses of these issues, often through pipeline or feasibility analyses or piloting of procedures, will ensure that the trial is uniformly conducted across study sites. The goal is to mitigate differences and to reduce the need to control for these issues during the analytic stage. The issues to be addressed in the protocol are:

Rationale for the Study. Every trial has a purpose that should be specified in study aims and hypotheses. The rationale provides a guidebook as to why the trial is necessary and what we will learn from the trial.

Theoretical Foundation and How it Informs the Innovation. Trials are essentially about innovations, either an introduction of a new idea or a change in an existing procedure or process. The innovation should be guided by theory which should help explain why the new or revised procedure or process is likely to result in improvements in the desired outcomes. And, the theory should be articulated into the procedures that will be tested. It is important to ensure that each component of the procedure is theoretically based to provide an explanation of why the component is needed as part of the innovation.

Defining the Study Population. Based on the study aims, the nature of the study population should be defined. The criteria for inclusion and exclusion should be specified to ensure that the innovation is reaching the appropriate target population. Generally, inclusion criteria relate to traits of the individual or setting. But, in the real-world criminal justice settings, other selection forces may be present that may be a function of the process or environment. Researchers may not have access to the full base population, which can potentially bias results. For example, in recruiting at a prison or jail, wardens or correctional officers may decide that offenders under certain security restrictions are ineligible for study participation. Selection

forces can vary between study sites because of factors that are often not evident. Also, sites may operationalize selection criteria differently. A recent article on multisite trials in health services research by Weinberger et al. recommended that the lead center develop selection criteria that allow each site to operationally define these criteria given the local environment (Weinberger et al. 2001).

Defining the Intervention. The innovation that is being tested should, by definition, be different than traditional practice. As discussed above, the intervention needs to be outlined in detail, including theoretical rationale, dosage and duration, procedures essential to implementation, and core components that will be tracked using fidelity measures (Bond et al. 2000). Most interventions consist of different components that are combined together to produce the innovation. It is also important to specify who will implement the intervention (the provider) and how they will become proficient in the skills needed to deliver the intervention. In an experiment, it is important to clearly delineate how the innovation will be integrated into the existing infrastructure of an organization and what potential issues are expected to occur and how these will be handled.

Defining the Control Group(s) or Practice As Usual. The main purpose of the control group in an RCT is to have a contrast to the innovation. The control group generally represents practice as usual, and allows the ability to distinguish subject outcomes. The control group is expected to have the same distribution of other factors that could potentially affect outcomes as the treatment group, and thus, avoids the possibility of confounding. To achieve this goal, randomization gives each eligible subject an equal chance of being assigned to either group in a two-armed design. Blinding, in conjunction with randomization, ensures equal treatment of study groups. In general, the services received by the control group should follow the principle of equipoise, meaning that there should be uncertainty about whether the control or the experimental condition is a better option (Djulbegovic et al. 2003).

Measurement and Observations. Another important component of a study is how effects of the intervention will be determined, i.e., what instruments will be used to measure outcomes and how and when these instruments will be used. Dennis outlines four types of variables to be collected: design variables (sites, study condition, weights, etc.), other covariates (demographics, diagnoses, etc.), pre- and post-randomization intervention exposure, and pre- and post-dependent variables (Dennis et al. 2000). Most experiments have multiple waves of data collection where the same data is collected at baseline and at specified follow-up points. Outcome measurements can be subject to a number of different biases, including recall bias for self-report measures, sensitivity and specificity issues (biomedical tests), and rater error (file reviews). The types of data to be collected should be based on the hypotheses to be tested and the availability and cost of data. For the main outcomes, it is often recommended to have multiple sources of measurement, such as self-report and urine-testing for drug outcomes. These can be put together to create combined measures (Dennis et al. 2004).

Quantitative Analysis. The analysis of multisite trial data is dependent on the types of measures collected and the format of the outcome and independent variables. While the trial is ongoing, interim analyses may be conducted, mainly for data monitoring purposes. Some trials in the justice system may require oversight from a data and safety monitoring board (DSMB) who will run reports on a regular basis to determine if subject flow and follow-up are adequate and if treatment effects are observed in the data. These interim data runs often use t-tests, ANOVAs, or effect sizes (Cohen 1988). For more formal analyses, multivariate models are used to account for the repeated measures and multisite designs. With repeated measures data, hierarchical models can be used, which nest observations within person over

time (Raudenbush 2001). Different types of models are available including latent growth models (LGM), which captures individual differences in change trajectories over time (Meredith and Tisak 1990). With multiple assessment points, LGM tests for both linear and nonlinear growth through the use of specified growth functions and additional growth factors (e.g., quadratic, piecewise linear models). LGM also allows for an examination of what factors predict changes in outcomes over time, controlling for both measured and unmeasured stable individual-level characteristics (Bollen and Curran 2006).

To an extent, a multisite trial allows for the ability to pool the overall impacts of treatment and the specific impacts of treatment within separate sites in the context of a single statistical model. The sites in a multicenter trial can be seen in this context as building blocks that can be combined in different ways by the researchers. The separate randomization procedures allow each center to be analyzed as a separate experiment as well as combined into an overall experimental evaluation where the researcher is able to identify direct and interaction effects in a statistically powerful experimental context. If the researcher was to identify subjects from all sites and then randomly allocate subjects to treatment and control conditions (without reference to site), then any statistical analysis of the impacts of site on outcomes would be nonexperimental and subject to the same threats to internal validity due to omitted variables common in nonexperimental research (see Smith 1990).

There is a cautionary note in analyzing multisite studies. A multisite study is not really a multicenter study where there is a random sample of sites themselves. Without a random sample of sites, a number of restrictions must be placed on the generalizations that can be made and the type of statistical models that can be estimated. Site must be defined as a fixed factor in the statistical models employed, because of the possible variation in the context of the sites. It is important to consider that the generalizability of these effects is limited and cannot be applied to the larger population. In the mixed effects model, a hypothesis about the average treatment effect of the population refers to the population of sites from which the sample of sites is drawn.

Weisburd and Taxman (2000) recommend a statistical process for using mixed effects model in analyzing multisite studies. A model for this example is presented in (25.1) below. This provides a technique for dealing with the multisite issues.

$$Y = \mu + S_i + T_j + ST_{ij} + O_{k(i,j)} + e_{k(i,j)} \tag{25.1}$$

where T_j is the larger population of centers for the mixed effects model; S_i is the average impact of a site (or center) on outcomes, averaged over treatments; ST_{ij} is the interaction effect between treatment and site, once the average impacts of site variation and treatment variation have been taken into account; and $O_{k(i,j)}$ is nested within a site by treatment combination.

TOOLS TO MONITOR IMPLEMENTATION: CONSORT FLOW CHARTS IN CJ SETTINGS

Understanding the flow of subjects through the different phases of an experiment is an important task to manage the quality and integrity of the experiment. In multisite trials, subject flow may be influenced by facets of the particular criminal justice system, the behavior of study subjects, or factors that are difficult to identify. In 2001, the Consolidated Standards of Reporting Trials (CONSORT) statement was issued as a way to improve the reporting of information from trials (Moher et al. 2001). The CONSORT flow chart illustrates how to

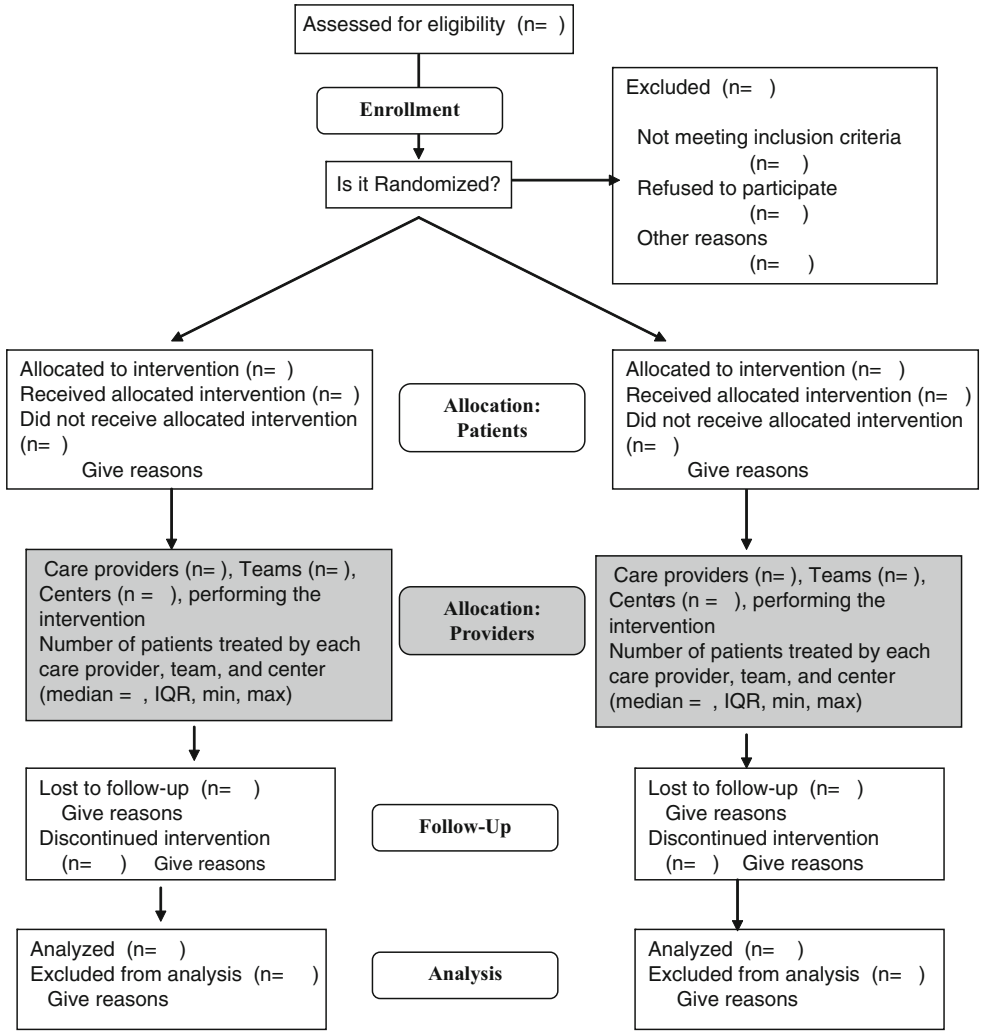

FIGURE 25.1. Sample consort flowchart for nonpharmacologic treatments.

report subjects through each trial phase; the flow chart was amended in 2008 to specifically address issues related to nonpharmacologic treatments (such as different types of therapy or different modes of sanctions). A copy of the amended chart is in Fig. 25.1 (Boutron et al. 2008). The CONSORT chart provides documentation for each phase of the experiment, and provides sufficient information to assess the quality of the implementation of the experiment.

This section describes the CONSORT flow chart for nonpharmacologic treatments (Fig. 25.1), as it relates to implementing trials in criminal justice settings. The unique challenges presented by these settings include not only the target population as potentially vulnerable to coercion and exploitation, but also reflects the difficulties of work in settings that are, by their nature, authoritarian and often not amenable to the rigors of randomized trials.

The first section concerns eligibility. Most trials involve decisions about specific inclusion and exclusion criteria that should be in place for the trial so that the results can be generalizable to the appropriate population. The identification of people that fit the desired criteria may involve a multistep process. For example, a specific protocol may require the exclusion of sex offenders. This will require the research project to define the term "sex offenders" and then develop specific criteria to meet this term. In practice, researchers will generally need to conduct a prior chart review of offenders to make a determination of which offenders might meet the eligibility criteria. This screening often occurs before the research interview or consent is given. Other types of frequent eligibility criteria are offense specific, geographical areas, age or gender issues, language, or other characteristics. In general, screening criteria should be geared towards obtaining a study population that would be appropriate for the intervention that will be tested. The question for the CONSORT chart is how many subjects meet the criteria, and what process is used to identify the target population. This defines the intent to treat group.

Prior to randomization, subjects may be dropped from the study because they fail to meet screening criteria or refuse to participate in the study. Once they are randomized, they may either receive their allocated intervention (control or treatment) or they may not receive it for a variety of reasons including not meeting the eligibility criteria when they are interviewed, dropping out of the study, transferring to another CJ facility, or having a change in criminal justice status. All of these are examples of postchart review factors that may affect participation in a study. The randomization process is such that once a subject is assigned to a condition, they are eligible for follow-up even if they only receive a partial dosage of the intervention or they do not complete the period of time devoted to the intervention.

The provider allocation section of the CONSORT chart was added in 2008 and reflects the need to characterize the settings where the treatment(s) take place. The researcher needs to identify all of the potential settings that might be involved in the experiment. For example, the range of settings could be treatment, criminal justice, courts, or other areas. To be documented is the number of providers, teams, and/or centers and the number of subjects in each area. In criminal justice settings, it will often be the specific organization that is of interest such as the parole office or prison where the trial may be implemented. Each organization will have different characteristics, and these require documentation.

A key issue is the follow-up rates for subjects in all arms of the study. The documentation should include the number that have been followed-up and the reasons for missed follow-ups. Reasons for attrition from a study are critical to ensure the study has sufficient power for the analyses. It is important to document follow-up rates at each wave, along with reasons for exclusions of any cases.

The last component of the CONSORT documentation is the result from the final outcome variables that relate to the study aims and the timing of all variables. The researcher is also required to document any assumptions that were made about the data.

Some criticisms have been made of the lack of detail present in the CONSORT chart (Mayo-Wilson 2007), specifically the lack of information on implementation details. As our case studies will demonstrate, the CONSORT chart does not capture a number of issues that can present threats to the internal validity of a multisite trial. Other scales and checklists have been developed to measure the quality of randomized trials, including the Balas Scale for health services research (Balas et al. 1995) and the Downs scale for public health research (Downs and Black 1998). A review of these scales found little widespread use of them and also found that most had not been tested for construct validity or internal consistency (Olivo et al. 2008).

CASE STUDIES OF RCTS IN CRIMINAL JUSTICE SETTINGS

The following sections present case studies of two RCTs that were implemented in criminal justice settings as part of the Criminal Justice Drug Abuse Treatment Studies (CJ-DATS1), a 10-center research cooperative sponsored by the National Institute on Drug Abuse from 2002 to 2008. Part of the objective of CJ-DATS was to use rigorous scientific methods to conduct studies in criminal justice settings. In all, six randomized trials were done in CJ-DATS. The purpose of the studies was to conduct studies at the point where offenders transit from incarceration to the community, referred to as the structured release or reentry process. The studies involved numerous agencies including prison, probation/parole agencies, prison or community treatment providers, and nonprofit agencies.

The case studies are presented to illustrate the issues regarding study design and management. We use as the conceptual framework for this analysis the threats to internal validity that are specified in many textbooks on conducting social experiments (Campbell and Stanley 1966). By examining these issues, as well as looking at how the researchers managed the study, we can learn about the issues inherent in implementing multisite trials.

THREATS TO INTERNAL VALIDITY

Confounding. The degree to which a third variable, which is related to the manipulated variable, is detected. The presence of spurious relations might result in a rival hypothesis to the original causal hypothesis that the researcher may develop.

Selection (bias). Both researchers and study participants bring to the experiment many different characteristics. Some of these characteristics, which may not have been included in the eligibility criteria, may influence those who participate in an experiment. These factors may influence the observed outcomes. The subjects in both groups may not be alike in regard to independent variables.

History. Study participants may be influenced by events outside of the experiment either at the onset of the study or between measurement periods. These events may influence attitudes and behaviors in such a way that it might be detectable.

Maturation. During the experiment, subjects may change and this change may result in how the study participants react to the dependent variable.

Repeated testing. Conducting multiple measures over periods of time may result in bias on the part of the study participants. These impacts may occur from the failure to recall correct answers, or the prospects of being tested.

Instrument change. An instrument used during the experimental period could be altered.

Mortality/differential attrition. Study dropouts may be the result of the person, the experiment, or the situation. An understanding of the reasons for the attrition and the differential rates of attrition can address the issue of whether the attrition is natural or due to some other factors.

Diffusion. The experimental category should be different than traditional practice, and movement from the experimental to control group or vice versa may pollute the assigned conditions.

Compensatory rivalry/resentful demoralization. The control group behavior could be tainted by the result of being in the study.

Experimenter bias. The researcher or research team may inadvertently affect the experiment by their actions.

1. *Step'N Out*. An RCT at six Parole Offices in five States

Step'N Out tested an integrated system of establishing target behaviors for offenders on parole, collaborative behavioral management (CBM). CBM requires the parole officers and treatment counselors to work together with offenders over a 12 week period, using a system of graduated sanctions and incentives. A more detailed description of the study can be found in Friedmann et al. (2008).

Theoretical Intervention. CBM has three major components. First, it explicitly articulates parole and treatment staff roles as well as the offenders' and the expectations of each party (role induction theory). Second, a behavioral contract defines the consequences if offenders meet or fail to meet those expectations. The behavioral contract specifies concrete target behaviors in which the offender is expected to engage on a weekly basis; these target behaviors include requirements of supervision and formal addiction treatment, and involvement in behaviors that compete with drug use (e.g., getting a job; enhancing nondrug social network) (behavioral targeting). Third, it regularly monitors adherence to the behavioral contract, and employs both reinforcers and sanctions to shape behavior. The motto is "Catching People Doing Things Right," which is to say, the intervention creates the conditions to notice and reward offenders for achieving incremental prosocial steps as part of normal supervision (contingency management). CBM establishes a systematic, standardized, and progressive approach to reinforcement and sanctioning to ensure consistency and fairness.

The CBM contract specifies expectations in terms of concrete target behaviors that the offender must meet before the next weekly session. Examples of target behaviors include producing a negative urine specimen; attending supervision and counseling sessions; and completing incremental steps toward getting a job or finding drug-free housing. These target behaviors are managed using a computer program, the Step'N Out COmputerized INput Environment (SNOCONE). The contract is printed out with copies for all three parties to sign and keep for their records. This process is completed weekly as part of standard parole conditions. The CBM contract is monitored weekly to expedite identification and reinforcement of compliance and sanction of noncompliance, and then the contract is renegotiated and printed for the following week. Compliance with the contract earns points and, when pre-established milestones are reached, material and social rewards.

Control Condition. Offenders were supervised by the assigned parole officer in the manner that is used by that jurisdiction. As shown in Table 25.1 below, each site had different standards for parole visits and drug testing based on standard practice.

Study Sites. Five CJ-DATS centers participated in Step'N Out – the University of Delaware (Wilmington parole office), Brown University/Lifespan Hospital (Providence parole office), UCLA (Portland, OR parole office), CT Department of MHAS (Bridgeport and Hartford parole offices), and George Mason University (Richmond parole office). All sites except Connecticut had one parole officer and one treatment counselor assigned to the CBM condition at their sites. In Connecticut, two or three officers were trained in CBM at each site, with one counselor also trained at each office.

Site Preparation. The initial two-and-half-day training for the Step'N Out teams occurred in December, 2004. This training brought together parole officer and addiction counselor teams and their supervisors. The training began with lecture presentation of the theoretical model and rationale for the intervention, research evidence for its components, and an outline of key elements. Training staff then demonstrated the key components of CBM. The remainder of the training focused on having the teams practice skills in case-based role plays with

TABLE 25.1. RCT implementation issues for Step'N Out

Issue	BP/HT	DE	OR	RI	VA
Place of randomization	Parole office	Halfway house	Parole office, switched to prison	Prison	Parole office
Selection bias	BP – 2 persons randomized but did not make initial parole session (CBM = 1, control = 1); HT – 2 persons (CBM = 1, control = 1)	85 offenders were not released to parole CBM = 42, control = 43	3 offenders not released (all control)	1 offender not released in CBM	Gang members were excluded
History	An event occurred (a parolee committed homicide) that resulted in a crackdown on parolees which increased violations			Recruitment was suspended in 2005. All subjects recruited were dropped from the study. Recruitment began again in July 2006 with offenders placed on electronic monitoring	
Maturation – time to touch base session	6.1 days (0–27 days) CBM = 5.9 vs. control = 6.2 for Bridgeport 7.7 days (2 to 25) for Hartford CBM = 7.4 vs. control = 12	68.3 days (0 to 470 days) CBM = 47 vs. control = 92.8	51.6 days (0 to 175 days) CBM = 51.2 vs. control = 52.1	17.1 days (5 to 69 days) CBM = 9.8 vs. control = 20.4	10.7 days (0 to 78 days) CBM = 10.4 vs. control = 10.9
Mortality/ differential attrition – 3 month FU	3M FU BP: CBM = 100% Control = 100% 3M FU HT: CBM = 100% Control = 95.7%	3M FU: CBM = 98.9% Control = 97.9%	3M FU: CBM = 79.4% Control = 87.5%	3M FU CBM = 100% Control = 85.7%	3M FU CBM = 79.5% Control = 91.3%
Mortality/ differential attrition – 9 month	9M FU BP: CBM = 97.2% Control = 97.2% 9M FU HT: CBM = 92.3% Control = 82.6%	9M FU BP: CBM = 97.2% Control = 97.2% 9M FU HT: CBM = 92.3% Control = 82.6%	9M FU: CBM = 73.5% Control = 90.6%	9M FU: CBM = 88.2% Control = 78.6%	9M FU: CBM = 64.1% Control = 78.3%
Selection-maturation interaction (average)	Control group saw parole officers once a month	Control group saw parole officers once a week	Control group saw parole officers two times a month	Control group saw parole officers once a week	Control group saw parole officers once a month
Diffusion		Possibly in group SA sessions since treatment sessions for control and CBM at parole office			Possibly in group SA sessions as treatment for both control and CBM was at parole office

(continued)

TABLE 25.1. (continued)

Issue	BP/HT	DE	OR	RI	VA
Compensatory rivalry		Anecdotal stories of resentment by control subjects of material and social rewards given to CBM			
Experimenter bias	Had a number of POs implementing CBM at each site; may have been differential implementation by PO. PO had few CBM clients, may have diluted effects	Changed CBM POs part way through the study, may have had different implementation styles	Changed CBM counselor part way through process, may have different implementation styles		CBM PO felt isolated, also had substantial non-CBM clients, may have diluted effects

reinforcement and corrective feedback. A checklist of the key elements for fidelity to the protocol guided the role plays and feedback. The teams were encouraged to negotiate roles with regards to initiating the role induction discussion, establishing goals and setting target behaviors, but the protocol recommends that the PO take primary responsibility for rewards and sanctioning and the counselor for problem-solving.

Additional on-site trainings were also scheduled due to the lag time between the initial training and the time that sites began recruitment, the addition of new sites, and staff turnover. A 2-day booster training session in September, 2006, focused on enhancing both the fidelity and finesse with which teams delivered the intervention.

Parole officers were also asked to tape random sessions. Feedback on the sessions was to be given on a monthly basis to build the skills of the officers in using the contingency management procedure.

Study Implementation. Table 25.2 outlines the issues encountered in implementing the study across the fives sites. One of the first challenges was the point of randomization. Because those who were randomized to the CBM condition had to be assigned to a specific parole officer (the one who had received the CBM training), randomization had to be done at a point where parole officer assignment had not yet been done. This point varied from site to site, with some assigning officers prior to an offender's release from prison, while others assigned the officer once the offender appeared at the parole office. While the ideal situation was to randomize at the parole office where the intervention was occurring, this was not possible at all sites, particularly in Delaware and Rhode Island. At those two sites, the agency practice was to assign the officer prior to an offender's release. In these sites, screening and randomization were done while the offender was still incarcerated; this allowed the offender who was randomized to CBM to be assigned to the CBM officer. In Oregon, randomization began at the parole office, but recruitment was very slow in the initial months of the study and it was determined that the research staff had a better working relationship at the prisons; thus, recruitment was moved to the prison after 6 months of commencing the experiment. In the end, half of the sites were randomizing in prison and half were randomizing at the parole office.

TABLE 25.2. RCT implementation issues in HIV/HEPC study

Issue	DE	KY	VA
Place of randomization	Halfway house	Prisons (2 male, 1 female)	Local jail
Start of follow-up	Randomization	Release date	Release date
Selection bias	Unknown selection forces for initial groups – had 406 randomized but only 391 released (96.3%)	Unknown selection forces for initial groups – had 184 randomized but only 125 released (67.9%)	Unknown selection forces for initial groups – had 113 randomized but only 67 released (59.2%)
History	A confidentiality breach in DE on a different project shut down interviewing and recruiting until new processes were put in place	KY was unable to recruit in the men's prisons	Jail staffing issues and changes in security procedures affected recruitment; suspended for 2 months
Maturation		Longer time between randomization and follow-up than DE because of release dates	Longer time between randomization and follow up than DE because of release dates
Repeated testing			
Mortality/Differential Attrition	30 Day FU: Std = 98%, NIDA = 100%, DVD = 99%; 90 Day FU: Std = 98%, NIDA = 98%, DVD = 96%	30 Day FU: Std = 100%, NIDA = 76%, DVD = 92%; 90 Day FU: Std = 100%, NIDA = 85%, DVD = 92%	30 Day FU: Std = 90%, NIDA = 73%, DVD = 94%; 90 Day FU: Std = 75%, NIDA = 60%, DVD = 88%
Selection–maturation interaction			
Diffusion	All arms of the study were all housed in the same facility for a period of time. Those in the intervention arms could share information with those in the control arms	All arms of the study were all housed in the same facility for a period of time. Those in the intervention arms could share information with those in the control arms	All arms of the study were all housed in the same facility for a period of time. Those in the intervention arms could share information with those in the control arms
Experimenter bias	Changed DVD interventionist part way through the study, also changed NIDA standard interventionist. Occasionally had same person do both arms		

A consequence of the place of assignment in these studies at the "transition" point is that not all randomly assigned subjects were exposed to the intervention. A substantial portion of those who were randomized in prison never made it to the parole office. In Rhode Island, this was due to the fact that many offenders "flattened" in prison, or served out their parole sentences in prison, because of the unavailability of space in halfway houses where offenders were supposed to go upon release. In Delaware, many offenders committed violations in the halfway house that resulted in them being returned to prison before they could begin their parole sentence. This group of "postrandomization dropouts" was problematic, because they were technically part of the study population to be followed, yet they were

not exposed to the intervention or any study activities. Many had no time in the community, postrandomization. Rhode Island dropped their initial cases and changed their recruiting practices so that all of their cases were to be placed on electronic monitoring after release. Delaware also changed their recruiting to make their screening closer to the initial parole session. These measures drastically reduced postrandomization dropouts. The final CONSORT chart for Step'N Out is given in Fig. 25.2 (note that when the study was done, the revised CONSORT for nonpharmacological interventions had not yet been published). There were a total of 93 postrandomization dropouts, with the majority ($n = 85$) from the Delaware site.

Having randomization occur in two different places (depending on the sties) meant that this needed to be accounted for in the analysis. The place of randomization has to be controlled

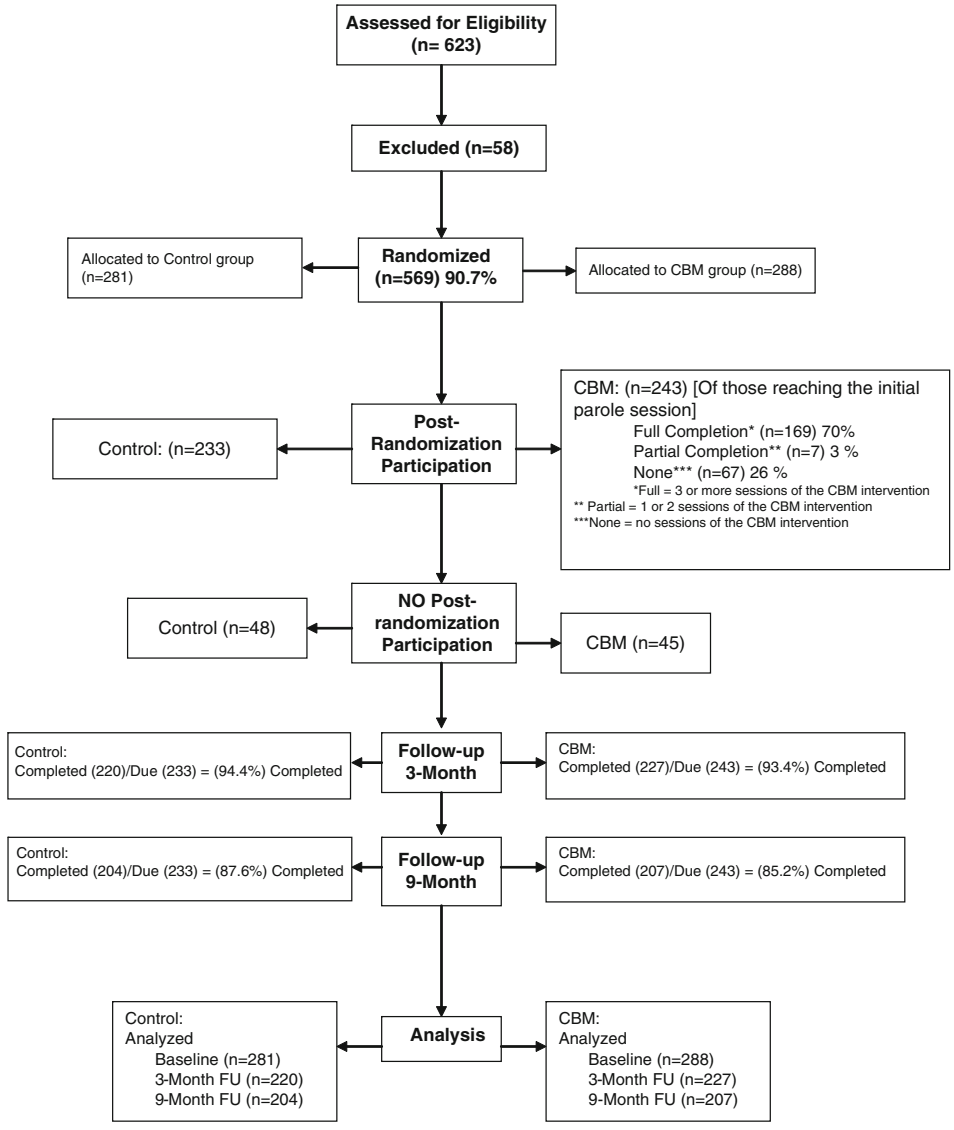

FIGURE 25.2. CONSORT flowchart of participants in Step'N Out study.

for when running statistical analyses and differences have been found in outcomes, according to this variable (Rhodes et al. 2009). Also, as is shown in Table 25.1, other selection forces also occurred during the study at different sites. In Virginia, gang members were never sent to the research interviewer for screening because a special program for gang members was being conducted with an assigned parole officer.

Another important site difference in implementation was the time it took for a person to begin the intervention after randomization. In some of the sites, those in the CBM group were able to begin parole services sooner than those in the control group, perhaps, due to the availability of a dedicated parole officer for the CBM intervention who did not have the same large caseloads as other officers. In both Delaware and Rhodes Island, it took about half the time for those in CBM to begin parole services as compared to those in the control group. This could lead to instrumentation issues, since the results from CBM, such as positive results, could indicate simply that those who begin parole sooner have different outcomes than those who started later (MacKenzie et al. 1999), as better outcomes could be a function of the timing rather than the superiority of the CBM intervention attributes.

Timing of instrumentation also affected the study. The follow-up clock for Step'N Out began with the date of the first parole session (the "touch base" session). Because the time to the touch base session was longer for some subjects than for others, the time to the follow-up interviews also varied. In some cases, the 3 month follow-up, occurred about 3 months after randomization (when randomization occurred close to the first parole session), but in other cases, there was over a year of time between randomization and the first parole session resulting in over 15 months between baseline and the 3 month follow-up. While the follow-up was generally capturing the first 3 months on parole, there was a wide variation in how much time had elapsed since randomization for each subject. It was also complicated since some sites (Delaware) also tended to have differential timing between follow-up periods.

An attrition-related issue involved missed follow-ups. The Delaware center, which had the largest number of clients in Step'N Out, had some staffing issues during the project and missed the 3 month follow-up window for a number of subjects. Because the main outcomes relied on data from the 3 month interview, provisions were made to obtain these data at a later date. Basically, at the 9 month interview, a full timeline follow back was completed which included all time back to the initial parole session. The prospects for historical and maturation issues increased by this procedure, especially with the different timeframes for data collection across the various sites. The strategy allowed for the study to obtain needed outcome data, but it also created potential recall issues, where subjects were asked to recall drug use, arrest, crime, and living situation data on a daily basis up to 18 months after it occurred. The varying lengths of the follow-up period may have affected the integrity of the data.

An issue that was inherent to the specific criminal justice systems in each state was the status of the control group. Because this group received parole and treatment services "as usual" in each state, the services received by the control group were not uniform across all sites, as shown Table 25.1. In some sites, parole officers met with subjects once a month; in others, once a week. These were due to standard conditions of parole but other variations could be due to the individual offenders who have different conditions which require different schedules for parole. The same issues applied to substance abuse treatment. While the screening instrument used to assess eligibility ensured that those in the study had substance abuse dependency issues, it was not necessarily required that members of the control group would receive treatment since this decision depended upon the parole recommendations or was a decision of the local parole officer. Not all parole officers recommended their clients

for treatment. Thus, variations occurred in the conditions under which the control group was supervised both within site and across sites.

2. The HIV/HEPC Study: A 3-Armed RCT at three Centers

This RCT was designed to test the efficacy of a criminal justice based brief HIV/HCV intervention administered during the reentry period of incarceration. Rates of both HIV and Hepatitis C are disproportionately high for those involved in the criminal justice system (Hammett et al. 2002), and the reentry period has been found to be a particularly risky time for offenders (Arriola and Braithwaite 2008). This study was a three-group randomized design, with (1) current practice, a group of offenders who saw an HIV awareness video shown as part of group sessions in the facilities during the pre-release process; (2) the NIDA Standard Version 3 HIV/HepC prevention intervention delivered by a health educator via cue cards in a one-on-one didactic setting and (3) the CJ-DATS Targeted, a near-peer facilitated intervention that used an interactive DVD format with gender/race congruent testimonials, also in one-on-one setting. Participants in all three groups were offered HIV and HCV testing. The full study design has been described elsewhere (Inciardi et al. 2007). The study was conducted by three CJ-DATS centers, the University of Delaware (Lead Center), the University of Kentucky, and George Mason University. The main purpose of the study was to test the interactive DVD method to determine if those who were exposed to a gender and culturally specific message had greater reductions in risky behaviors than those that receive the NIDA standard and current practice arms of the study. Follow-ups for this study were done at 30 and 90 days after either randomization or the date of release from the prison.

Theoretically Driven Intervention. Based on focus groups with offenders, the study team developed a DVD that addressed issues related to reducing risky behaviors that involve the transmission of HIV/Hep C. The messages were then developed to be consistent with the gender and ethnicity (i.e. Caucasian, African-American) of the offender. The theory was that the same messages delivered in a gender-culturally specific manner would result in greater compliance. The DVD arm of the study was administered by a near-peer, usually a former addict and/or someone with HIV or HCV. There were four separate DVDs, based on race/gender combinations (black male, black female, white male, white female) and they included testimonials from former offenders and persons infected with HIV and HCV discussing their situations. Peer facilitators have been found to be effective in reducing HIV risk behaviors in needle sharing populations (Latkin 1998).

Control Conditions. In this study, two control conditions were used: (1) traditional practices, where a group of offenders are shown a HIV awareness video as part of group sessions in the facilities in the prerelease process; and, (2) the NIDA Standard Version 3 HIV/HepC prevention intervention delivered by a health educator via cue cards in a one-on-one didactic setting (National Institute on Drug Abuse 2000). These represent two different conditions that reflect the variety of practice that might be used.

Study Recruitment. Potential participants were shown a video on HIV prevention in a group setting while still incarcerated (see section on setting by site), and were given information about the study and asked if they would like to participate. If they said yes, they were then screened individually to determine eligibility, and if eligible, they went through the informed consent process and were randomized into one of the three arms of the study. Testing for HIV and HCV was done at baseline and was voluntary and did not affect participation in the rest of the study. The main point of the study was to provide information on HIV and HCV prevention to offenders just prior to their release.

Implementation. Table 25.2 presents some of the implementation issues of this study. Most of the challenges resulted from the varied settings where the intervention took place. For Delaware, the study was implemented in a halfway house facility, a place which ensured an adequate flow of subjects and continuous access to subjects. This site selection was convenient but in many ways altered the original intent of the study, as the offenders at this facility were already at risk in the community since they reside in the halfway house but still can spend time in the community. In Kentucky, the study was implemented at a men's prison and a women's prison, while in Virginia, it was implemented at a local jail. Because the point of the follow-up period was to obtain information on risky behaviors in the community, the start of the follow-up clock was different according to site. For Delaware, because subjects were already at risk at the time of randomization, the follow-up clock started at the randomization date. In both Kentucky and Virginia, the follow-up clock did not start until the subject was released from the institution. This situation created a number of problems including the fact that the type of site (prison, jail, halfway house) was completely nested within state, and that the time at risk was not comparable among all subjects. Those in Delaware may have had some time at risk prior to being enrolled in the study. While baseline behavior data were gathered for the period prior to the last incarceration, the follow-up data collected may not reflect the same period of initial risk for those in Delaware as for those just released from prison and jail in Kentucky and Virginia.

Another issue that may have affected implementation was the initial selection of the group of offenders to watch the video at each site. Generally, these groups were gathered by personnel at the facilities who were given the criteria for the study (offenders had to be within 60 days of release, speak English, and not have cognitive impairments). Researchers were not often privy to the decisions about the type of offenders that participated in the sessions. It may be that different types of groups were selected for inclusion at the different sites.

Since follow-up was dependent on release in both Kentucky and Virginia, a substantial percentage of those randomized were ineligible for follow-up because they were not released during the time frame of the study. While those randomized were initially scheduled for release within 60 days of randomization, the release date often changed because of pending charges or other issues that occurred in the facility. In Kentucky, 33% of the randomized population was never released and in Virginia, 41% of the randomized population was not released during the timeframe of the study. This issue had the potential to create selection bias and differential attrition. The final CONSORT chart for this study is given in Fig. 25.3. Note that a good number of those assessed were lost both prior to completing the baseline and also postrandomization, because of prison release issues.

Finally, in one site, the same interventionist, a peer counselor, administered the NIDA arm and DVD arm for a period of time. Bias should have been minimized, since the interventionists could also do the testing for those in the control arm. There is a potential that the interventionists provided information to that group that was specific to their arm of the study. The interventionists were not blind to the study allocation of the subjects.

DISCUSSION

Multisite trials are important tools to accomplish scientific goals of (1) testing the efficacy of an innovation; (2) simultaneously "replicating" the study in multiple sites to determine the robustness of the innovation; and (3) allowing the unique characteristics of agencies,

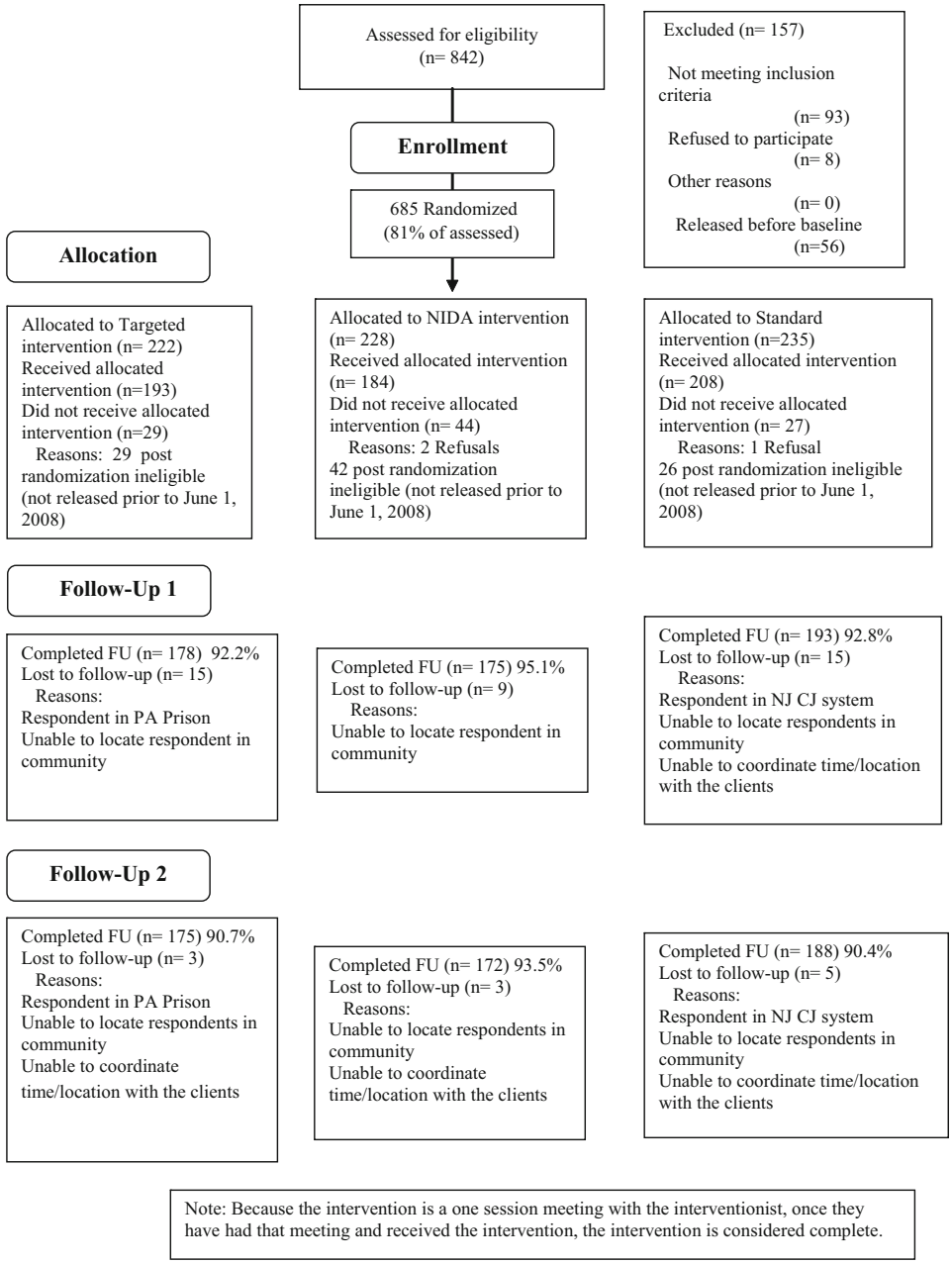

FIGURE 25.3. CONSORT flowchart of participants in HIV/HEPC study.

communities, and/or target populations to be concurrently examined in the test of an innovation to specify acceptable parameters. The trials represent the science of conducting field experiments but the process also reveals that there is an art in the process of conducting randomized trials. It is the art – the crafting of the experiment, the mechanisms to conduct

the experiment, and the daily decisions regarding the handling of "real world" situations – that determines whether the experiment achieved the goal of testing the theoretically based innovation.

The analysis in this paper presented the art in terms of issues that have been identified as threats to internal validity, threats also identified by others (Dennis et al. 2000). We used these as tests of internal validity to demonstrate how experimental processes and research designs affect the integrity of an experiment, and the findings from the experiment. These two experiments allow us to draw lessons about study management to reduce the degree to which the threats affect the dependent variables:

1. *Confounding*. An important step in reducing the influence of third variables is to conduct three pipeline flows: (1) the research process in each participating site; (2) the core components of the innovation should be outlined to determine what other factors affect the independent variable; and (3) the potential mediators or moderators that may affect the independent variable. All three of these pre-experimental processes are important to minimize extraneous variables. The CONSORT flow chart with the suggested new modifications provides a tool that is useful to complete at each site to raise questions about the influence of other factors or variables. That is, the researcher might consider the CONSORT flow chart as an implementation chart that can be used at all stages of the experiment to proactively identify potential issues that the researchers should explore and examine how they will impact study processes and findings.
2. *Selection Bias*. The inclusion and exclusion criteria of an experiment are critical in assessing potential sources of bias. But, more important than simply analyzing these a priori decisions, is to map these criteria to the study population to ensure that the research processes are not enhancing or affecting selection bias. For example, in the Step'N Out experiment, the researchers in one study site were unaware that the agency did not release offenders without the use of extra controls (e.g., electronic monitoring). Yet, in the early stages of the experiment, when this was discovered, it resulted in consideration of new eligibility criteria. Maps of the system would assist in this process. The two multisite trials analyzed here had other selection biases that crept into the study such as the use of different types of recruitment sites in the HIV/HEP C study. While this variation was useful in assessing the elasticity of the innovation across different correctional settings (prison, work release, jail), it introduced new problems, including a large population of postrandomization ineligibles, who were not released from the prison/jail within the study period of 12 months. This created a potential bias because it affects the intent to treat group, and the differences in releases versus detainees must be assessed to determine the degree to which the study protocol affected study outcomes. The use of a historical control group can mitigate issues with selection bias for this part of the population and can also reduce other biases, including diffusion, compensatory rivalry, and experimenter bias. In this case, the researcher must usually rely on records to obtain data for the historical group, but this strategy is useful, especially when the experimental condition appears to be clearly superior to the treatment as usual and denying it to a portion of the current population would be unethical. Another related issue is to have a historical control group that could define the base rates for the dependent variables. A historical control group can then help determine whether the inclusion and exclusion criteria affect the dependent variable(s).

3. *History*. The impact of events outside of the experiment, either at the onset of the study or between measurement periods, may influence attitudes and behaviors in such a way that it might be detectable. In Step'N Out, a crackdown on parolees in Connecticut occurred during the follow-up period, resulting in a higher number of parole violations than would have otherwise occurred. While it was felt that this event did not differentially affect the study groups, it was presumed that the overall number of parole violations was higher in the study than it normally would have been. This information is important to know and to report with all analyses of the study, especially as parole violations were considered a main outcome and ended up being much lower for the CBM group than the control group. It is important to be aware of events that occur that may affect the study outcomes.
4. *Maturation*. During the experiment, subjects may change and this change may result in how the study participant reacts to the dependent variable. In both studies, some sites had longer times between randomization and the follow-ups because of timing and release issues and some subjects also had longer times until their follow-ups were completed. In these cases, it is difficult to determine if the intervention caused positive changes or if more time in the community and exposure to other factors could be a possible cause. Future studies could focus on ensuring that randomization and followup schedules are implemented so that there is little room for variance and that research interviewers are trained to understand the importance of completing followups in the given time windows. For example, studies that focus on the release period could delay randomization until an offender is released to avoid the uncertainty associated with release dates.
5. *Mortality/differential attrition*. Study drop-outs may be the result of the person, the experiment, or the situation. In a couple of sites in Step'N Out, the CBM group had lower follow-up rates at both 3 and 9 months, indicating that offenders may have found it more difficult to complete the CBM intervention in these sites and to stay connected with the parole office. The implementation of the intervention at these sites should be explored to determine how it differed from other sites. In Oregon, for example, those participating in CBM, had to attend parole sessions at a different office, located at the treatment center, which may have been inconvenient for them. In another site, the CBM officer was required to also maintain a caseload of regular parole clients, and she often had scheduling conflicts with her CBM clients. These issues provide insight into how implementation can affect participation. One suggestion for future iterations of Step'N Out was to have it be an office-based, rather than officer-based intervention, where the entire parole office practiced CBM.
6. *Diffusion*. The experimental category should be different than traditional practice and movement from the experimental to control group or vice versa may pollute the assigned conditions. While there was no known crossover from one study condition to another, there was potential in both studies for offenders from different conditions to mix with each other and share information. In Step'N Out, one site had a control group that was not extremely dissimilar to the CBM group, as the study was implemented at a parole site where specialized parole services were already offered and officers met with their clients weekly. These officers often used novel methods of sanctions and rewards. In these cases, it is difficult to determine the explicit effects of the intervention under study, as: (1) the control group may have been influenced by the intervention and (2) the control group may not reflect current practice accurately. For multisite trials, this variation among current practice at sites is an issue that is

often raised. An initial feasibility study at each site should identify potential issues and provide insights on how each site works. Each study protocol should have some flexibility so that it can be adapted to multiple sites, but also must specify minimum conditions for the control group that must be met in order for the site to participate.

7. *Compensatory rivalry/resentful demoralization*. The control group behavior could be tainted by the result of being in the study. This is more of a possibility in longer lasting interventions where the treatment under study appears as a "better" alternative to those who are targeted for treatment. In Step'N Out, there was some evidence that persons wanted to be in the CBM condition (often asking for this at randomization) and felt those in CBM were getting an "easy ride." This could lead to resentment from the control group, especially if they were aware that the CBM group would receive material rewards for good behavior and they would not. This could create a disincentive for good behavior for those in the control group and study results could be biased upward, with results better than they should be under normal circumstances. Researchers should also be careful in using language to describe the study conditions to ensure that the experimental condition is not seen as highly preferable to the current practice and should also work with staff at the agencies to train them not to present the intervention as an improvement or something better, as staff attitudes are usually quickly conveyed to offenders.
8. *Experimenter bias*. The researcher or research team may inadvertently affect the experiment by their actions. Both of the studies we examined here had interventionists on-site at justice agencies who were not blinded to the study condition assignment of the offenders. Blinding cannot be done for the interventionist in these types of studies as they must know what condition a person is in so that they can deliver the correct intervention to them. While a number of measures were in place to ensure fidelity (tapes in Step'N Out, checklist in HIV), there were concerns that parole officers may have implemented CBM differently at different sites and even at the same site, based on the person involved. The same is true for the health educator and peer interventionist in HIV, especially the peer counselor who had less of a script to work with and used more of her own experience in working with her clients. While the tapes were a good idea in Step'N Out, they were not used effectively as a feedback tool, as there was a long delay in coding and providing feedback to the parole officers. Because delivery of the intervention is such a key component of the study, researchers should make provisions for ongoing and timely fidelity monitoring of the intervention, with a clear plan on how fidelity will be tracked, how feedback will be provided, and how corrective actions will be taken (Carroll et al. 2007).

The multisite trial also affords the opportunity to assess issues of external validity or the generalizability of the study. The simultaneous replication of an innovation in various settings, that are defined by sociopolitical environments that affect both the innovation and the research processes, allows the innovation to be tested for its elasticity or "fit" in various environments. It is these "tests" of the sociopolitical environments, not just the innovation itself, that demonstrate that the innovation has both internal and external validity. An innovation that cannot be implemented in a myriad of research settings is unlikely to be useful in practice, since research prototypes are seldom fully employed in practice. Alignment to an environment accounts for the acceptability of an innovation. It is important in experiments to identify the factors that affect innovation rejection or the characteristics or features when the organism is unable to be accepted by the existing body. The multisite trial offers the ability to assess these issues.

If, however, the political or organizational practices make it impossible to implement the intervention, no information can be gained from the effort (Rossi et al. 1999).

Another important area in trials, especially in justice settings, is the potential impact of the consent process on the study. Prisoners are considered a vulnerable population and while not all justice research is done with prisoners, all populations in this arena should be considered vulnerable and precautions should be taken to ensure human subjects protection. Research center Institutional Review Boards (IRBs), as well as IRB for justice agencies, should adhere to the following, as outlined by the Office of Human Research Protections (OHRP) in part 46 subpart C pertaining to prisoners in research:

- A majority of the IRB (exclusive of prisoner members) will have no association with the prison(s) involved, apart from their membership on the IRB.
- At least one member of the IRB must be a prisoner, or a prisoner representative with appropriate background and experience to serve in that capacity, except that where a particular research project is reviewed by more than one IRB, only one IRB has to satisfy this requirement.

When going through the informed consent process with offenders, the following points should be stressed (1) Participation or nonparticipation will not affect their legal status; (2) Individual or identifiable data collected in the study will not be made available to criminal justice authorities; (3) Choosing to participate or not participate will not affect their length of stay in the criminal justice system; (4) Their criminal justice records will not mention their choice to participate or not participate in a study protocol and this decision will not be a part of any official record on file. But this process can result in some selection bias in that some subjects may elect not to participate in the study based on the consenting process. This may or may not impact the study.

As a researcher considers validity threats to a multisite experiment, they should ask themselves the following five questions and to carefully consider how the decisions made affect the experiment:

1. How does the issue affect the flow of subjects on the CONSORT chart?
2. How does the issue affect inclusion and exclusion (selection forces)?
3. Does the issue disrupt the theory underlying the intervention? Or does it disrupt the treatment as usual?
4. How might the issue affect the dependent variable(s)?
5. How might the issue affect possible confounders?

CONCLUSION

Field experiments are well known to test an innovation as well as the environment where an innovation can be offered. In addition, the field experiment requires the researcher and/or research team to be sensitive to the impact of design and implementation decisions on the integrity of the experiment, as well as the generalizability to the wider target population. This article has been designed to help researchers think about the issues related to internal validity threats, and how decisions that are made about the target population, eligibility processes, intervention design and processes, and instruments. In addition, attention to attrition during waves of follow-up is critical to assess the quality of the experiment. Managing one study site presents sufficient challenges while managing more than one study site requires the

researchers and investigator(s) to be cognizant of how each decision might affect whether the goals of the study can be obtained. In both case studies, we have examples where practical realities crept into the experiment. More work will need to be done to determine the impact on the outcome variables. In the end, it is apparent that scientists must also be good research administrators to ensure the one or more study sites uphold the experimental design, and that each decision is weighed against the concern about the integrity of the design. Research requires an attention to methodological details from design to each phase of data collection to analysis. These details are the cornerstone of whether a study answers the questions it was designed to address.

Acknowledgements This study was funded under a cooperative agreement from the U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute on Drug Abuse (NIH/NIDA) to George Mason University (Grant U01 DA016213–01, Action Research to Advance Drug Treatment in the CJS). The funding for this cooperative agreement was supplemented by the Center for Substance Abuse Treatment, Bureau of Justice Assistance, Centers for Disease Control and Prevention, and National Institute on Alcohol Abuse and Alcoholism. The authors acknowledge the collaborative contributions by federal staff from NIDA and the other nine Research Center grantees of the NIH/NIDA CJ-DATS Cooperative. The contents are solely the responsibility of the authors and do not necessarily represent the official views of NIH/NIDA or other participants in CJ-DATS.

REFERENCES

Arriola KRJ, Braithwaite R (2008) Male prisoners and HIV prevention: a call for action ignored. Am J Public Health 98(9 Suppl):S145

Balas EA, Austin SM, Ewigman BG, Brown GD, Mitchell JA (1995) Methods of randomized controlled clinical trials in health services research. Med Care 33(7):687–699

Bollen KA, Curran PJ (2006). Latent curve models: a structural equation perspective. Wiley Interscience, Hoboken, NJ

Bond G, Evans L, Saylers M, Williams J, Kim H (2000) Measurement of fidelity in psychiatric rehabilitation. Ment Health Serv Res 2(2):75–87

Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P (2008) Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med 148(4):295–309

Campbell DT, Stanley JC (1966) Experimental and quasi-experimental designs for research. Houghton Mifflin, Boston

Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S (2007) A conceptual framework for implementation fidelity. Implementation Science 2:40–48

Chow SC, Liu JP (2004) Design and analysis of clinical trials, 2nd edn. Hoboken, NJ: Wiley Interscience

Cohen J (1988) Statistical power analysis for the social sciences, 2nd edn. New York: Lawrence Erlbaum Associates

Dennis ML, Perl HI, Huebner RB, McLellan AT (2000) Twenty-five strategies for improving the design, implementation and analysis of health services research related to alcohol and other drug abuse treatment. Addiction 95:S281–S308

Dennis M, Godley SH, Diamond G, Tims FM, Babor T, Donaldson J et al. (2004) The cannabis youth treatment (CYT) study: main findings from two randomized trials [Abstract]. J Subst Abuse Treat 27(3):197–213

Djulbegovic B, Cantor A, Clarke M (2003) The importance of the preservation of the ethical principle of equipoise in the design of clinical trials: relative impact of the methodological quality domains on the treatment effect in randomized controlled trials. Account Res: Policies Qual Assur 10(4):301–315

Downs S, Black N (1998) The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 52(6):377–384

Friedman L, Furberg C, DeMets D (1985). Fundamentals of clinical trials. PSG Publishing Company, Inc., Littleton, MA

Friedmann PD, Katz EC, Rhodes AG, Taxman FS, O'Connell DJ, Frisman LK, et al. (2008) Collaborative behavioral management for drug-involved parolees: rationale and design of the step'n out study. J Offender Rehabil 47(3;3): 290–318

Hammett TM, Harmon MP, Rhodes W (2002) The burden of infectious disease among inmates of and releasees from US correctional facilities, 1997. Am J Public Health 92(11):1789–1794

Inciardi JA, Surratt HL, Martin SS (2007) Developing a multimedia HIV and hepatitis intervention for drug-involved offenders reentering the community. Prison J 87(1):111–142

Latkin C (1998) Outreach in natural settings: the use of peer leaders for HIV prevention among injecting drug users' networks. Public Health Rep 113(Suppl 1):151

MacKenzie D, Browning K, SKroban S, Smith D (1999) The impact of probation on the criminal activities of offenders. J Res Crime Delinq 36(4):423–453

Mayo-Wilson E (2007) Reporting implementation in randomized trials: proposed additions to the consolidated standards of reporting trials statement. Am J Public Health 97(4):630–633

Meredith W, Tisak J (1990) Latent curve analysis. Psychometrika 47:47–67

Moher D, Schulz KF, Altman DG (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 357(9263):1191

National Institute on Drug Abuse (2000) The NIDA community-based outreach model: a manual to reduce the risk of HIV and other blood-borne infections in drug users No. (NIH Pub. No. 00-4812). National Institute on Drug Abuse, Rockville, MD

Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ (2008) Scales to assess the quality of randomized controlled trials: a systematic review. Phys Ther 88(2):156–175

Pocock S (1983) Clinical trials: a practical approach. Wiley, New York

Raudenbush SW (2001) Comparing personal trajectories and drawing causal inferences from longitudinal data. Annu Rev Psychol 52:501

Rhodes A, Taxman F, Rose JR, Friedmann P (2009) Rapport between parole officers and offenders: a mediation analysis. Unpublished manuscript

Rossi P, Freeman H, Lipsey MW (1999) Evaluation: a systematic approach, 6th edn. Sage Publications, Thousand Oaks, CA

Smith JK (1990) Alternative research paradigms and the problem of criteria. In E. G. Guba (Ed.), The paradigm dialog (pp. 167–187). Newbury Park, CA: Sage

Weinberger M, Oddone EZ, Henderson WG, Smith DM, Huey J, GiobbieHurder A, et al. (2001) Multisite randomized controlled trials in health services research: scientific challenges and operational issues. Med Care 39(6): 627–634

Weisburd D, Taxman FS (2000) Developing a multicenter randomized trial in criminology: The case of HIDTA. Journal of Quantitative Criminology 16(3):315–340

Weisburd D (2003) Ethical practice and evaluation of interventions in crime and justice: the moral imperative for randomized trials. Eval Rev 27(3):336–354