Baseline Assessment, Study Entry, and Stabilization: Double-Blind Clinical Trials in Drug Dependence

John Grabowski, Gila Arnoni, Ronith Elk, Howard Rhoades, and Joy Schmitz

INTRODUCTION

The conduct of clinical trials in psychopharmacology, including the area of drug dependence, has special complexity because the disorders reflect the interplay of pharmacological, biological, behavioral, and environmental determinants. Many issues concerning clinical trials in psychopharma-cology have been addressed by Prien and Robinson (1994). The study of new medications for the treatment of drug dependence has presented challenges (Blaine et al. 1994). In the case of cocaine dependence, these have included difficulty in recruiting uncomplicated patients (i.e., those who do not have multiple medical, psychiatric, or severe additional drug dependence problems), and high dropout rates. Many of the problems are not unique to drug-dependent patients generally, or cocaine-dependent patients specifically, although they are still described in these terms (Blaine et al. 1994). Rather, there are commonalties in problems, research issues, and probably treatment elements across disorders that are heavily imbued with both behavioral and biological components. This is particularly true with respect to issues of "compliance" or adherence to treatment. O'Brien and McLellan (1996) reported that observed rates of noncompliance with medication regimens and other features of treatment are equally common with disorders such as diabetes and heart disease, as they are in substance abuse. Clearly, differences among habitual behaviors (Levison et al. 1983), and medical disorders where problems of compliance are common, cannot be ignored, but the drug-dependent population is not unique with respect to adherence to treatment regimens. At least some of the issues can be resolved by precision in defining the design and mechanics of each clinical research study. Some clinic-specific efforts have been described previously (Elk et al. 1993).

Conducting definitive scientific studies of medications becomes more difficult in the face of noncompliance or multiple disorders. Yet

many of these difficulties can be confronted and resolved in planning systems for the conduct of the research (Kartzinel et al. 1994). Particularly important are the initial contact, intake, and period of stabilization before study entry. Overeagerness to conduct the clinical trial may produce simple procedural errors that negate the studies value. This can result in failure to demonstrate efficacy where it exists, or perhaps more damaging, produce reports of efficacy in the absence of actual benefit.

It is assumed at the outset that there is little value in nonblinded trials lacking placebo controls. Indeed, as Kupfer and colleagues (1994) have noted, "it can take years to overcome the results of flawed trials," and at least some of the research in this field has been devoted to that task. Thus, this chapter is pro forma and the information should be familiar. Nonetheless, investigators may benefit from rethinking standards, biases, preferences, and idiosyncrasies of the field. The goal is to describe some mechanical steps contributing to effective baseline assessment and study entry. Precision at this stage is critical insofar as subsequent measurements hinge on the validity of initial contact, intake, and stabilization. The procedures and issues considered here are drawn from experience and problems evident in the current literature. Strategies described have been used successfully over the years at the Substance Abuse-Medications Development Research Center (SARC), at the University of Texas. Some theoretical and practical issues in clinical trials that might influence the generalizability of study results are also discussed.

BASELINE ASSESSMENT

Standardization

Underlying the data collection process is the need for standardization of recruitment, baseline assessment, and stabilization. Standardization is critical if replicable and generalizable results are to be obtained from clinical trials evaluating new medications for the treatment of drug dependence. At the level of mechanics, important factors may be overlooked by novice and experienced investigators alike.

The SARC Clinic

The SARC Clinic was developed under National Institute on Drug Abuse (NIDA) demonstration grant DA 06143 as a new research treatment facility in 1989 and has no nonresearch service component.

The focus was on new treatments for drug dependence and reduction in HIV transmission. There existed the opportunity to explore optimal procedures (Elk et al. 1993), although there is continuing refinement. A major issue in developing the clinic was to avoid pitfalls including eliminating or minimizing common deterrents to patients seeking treatment or deflecting them from research participation. Many of these problems emerged during the initial contact and stabilization. There is benefit in detecting misassignments or other problems early, thereby increasing efficiency of the process.

An important feature in the development of the clinic was that it should be comfortable and provide readily accessible service to drug-dependent patients with little risk. It seemed likely that this would maximize the baseline level of retention with respect to obvious resolvable problems. The physical environment is well maintained and is a standard reasonably appointed outpatient clinical care facility resembling that of other specialty clinics. A variety of provisions to assure comfort, safety, and efficiency of service were also established, while at the same time maximizing collection and accuracy of data for clinical trials. Those devoted to assuring the safety and comfort of staff and subjects/patients are listed in table 1 (also see Grabowski et al. 1993).

The description derives from approximately 25 projects implemented between 1988 and 1994. These have involved about 1,000 enrolled patients and many more initial contacts and screens, in studies of opiate, cocaine, nicotine, benzodiazepines, and other forms of drug dependence. The studies, which included a range of special populations as well as “uncomplicated” patient population, included about 31,000 urine screens, 67,000 doses of medication, and multiple administrations per patient of Profile of Mood States (POMS), Addiction Severity Index (ASI), and other instruments to each patient.

Patient Recruitment and Advertising

Advertising is a common means to obtain subjects for clinical trials; each successive advertisement increases the number of telephone calls. Although not yet thoroughly documented, there appear to be differences

TABLE 1. *Examples of fixed clinic-wide contingencies and the nature of consequences for patients.*

1. Regular attendance for continued treatment +/-
2. Maintain appointment time for counseling +/-
3. Maintain appointment time for medication +/-
4. Complete data and information update forms +/-
5. Return medication bottles +/-
6. Provide urine samples for drug screens as scheduled +/-
7. Arrive and depart in reasonable time (no loitering) -
8. Maintain clean air (no smoking) -
9. Contribute to a physically healthy clinic (no weapons) -
10. Support the clinic as the sole vendor (no drug dealing) -
11. Responsiveness to chemistry laboratory findings (no arguing)

NOTE: This table lists issues/behaviors that underlie problems in some drug dependence treatment clinics. Focus on these issues often interferes with service delivery. Generic provisions can be added or eliminated as needed. Positive (+) and negative (-) consequences must be clearly stated and systematically applied. The goal is specification of positive consequences where the absence of that consequence is itself unpleasant. Items 9 and 10 have attached consequences of warnings and potential discharge. Some issues such as discussion of accuracy of laboratory drug screen results have neither positive nor negative consequences; they are not open for discussion just as blood pressure readings are medical test results accepted without discussion.

in the populations as a function of advertising site, even within the same newspaper (e.g., front news section, sports, entertainment sections). These differences also prevail as a function of contacts with emergency rooms, psychiatric facilities, and the extent to which current patients refer new patients. This affects the rate of acceptable patients for any particular study.

Screening

Sites that advertise for subject patients must screen call-in and walk-in candidates alike. The proportion of acceptable subjects depends in part on the specificity of the advertisements. Some individuals call because they have learned from friends, or from other treatment sites, that treatment research opportunities are available. Since treatment research sites often pay for initial interviews and other time devoted to research, it can be expected that some individuals call for the opportunity to earn money. The number of false-positives invited for a full-intake screening appointment depends in part on the adequacy of the prescreening interview whether administered by telephone or in person (table 2).

Two problematic features emerge in recruitment at this level. One is that different information accrues to subjects depending on the source (advertisements, professionals, friends). Second, there is a tendency to treat potential subjects who walk in differently from those who make initial contact by telephone. Using the "bird in the hand" philosophy, an investigator's eagerness to enroll subjects may lead to special provisions being made for subjects who are already at the site. The inherent bias in these differences dictate that all subjects should have the same initial screening interview, whether by telephone or in person. Until it can be demonstrated that there are no differences between the patient who takes the time and trouble to attend the clinic for initial screening and those who call in, differences should be assumed.

Each deviation risks additional variability. Following this constant first contact, the same procedures are applied to patients regardless of source of entry. There is no difference between subjects in scheduling of appointment; for example, candidates who call, are referred, or walk in are scheduled in the first available intake session. No preferential provisions are used in this regard.

The use of the prescreening form also permits the researcher to obtain rudimentary data on the characteristics of individuals from the community who are seeking treatment. This provides additional information for a cumulative database on the status of treatment research seekers in the community.

Some special prescreening provisions should exist when there are multiple investigators with multiple studies at a site. The "big primate" principle may prevail in assignment, with the most senior investigator

TABLE 2. *Prescreening (telephone or walk-in) questionnaire.*

A. Introduction

1. *Treatment Research Clinic, this is (your name). May I help you?* Determine why individual is calling. If not immediately obvious from response, ask:
   —*Are you calling about an advertisement?* (If so, which ad?)
   —*Are you calling about receiving treatment?* (If so, what program?)

2. Preamble
   We have several different research programs available to provide different treatment for different drug or medication-related problems.
   I must ask you some questions to learn if you qualify for these and to decide which one might be best for you.
   Before I do, I want you to know that all of the information you give me will be strictly confidential.
   I will not ask for your name or telephone until we complete the interview so that you can feel free to answer without any problems if you decide not to continue. If you decide to make an appointment, I get the necessary information. I will now ask the questions—may I start?

B. Determination of Study Type

1. *For what type of drug or medication-related problem are you seeking treatment?* (circle 1): None, Cocaine, Opiates, Anti-anxiety medications, nicotine, other_____
   If cocaine:
   a. *Are you currently having a problem with depression?* (Elaborate) YesNo
   b. If yes: *Do you think your depression is only a result of your cocaine use or does it seem to be a separate problem?* If response = separate, refer to Cocaine+Depression Project.

2. Circle: Male Female

3. If female: *Are you pregnant?* YesNo

(Use flowsheet to determine procedure for pregnancy study.)

TABLE 2. *Prescreening (telephone or walk-in) questionnaire (continued).*

4. *Have you have a positive TB skin test?* Yes No
   a. *If yes: Have you had treatment for it?* Yes No
   b. *If yes*: *How long was the treatment?* _____

C. Standard Questions (all studies)
   1. *How did you hear about us?*
      _____

   2. *How long have you been using this drug?*
      _____

   3. *Have you had any therapy for your drug use in the past 6 months?* _____

   4. *What is your ethnic background? a. Caucasian, b. Black, c. Hispanic, d. Asian, e. Other*

   5. *How old are you?* _____

   6. *What is your Zip code?* _____

D. GO TO APPROPRIATE STUDY-SPECIFIC QUESTIONNAIRE.

NOTE: All potential subjects are queried with the above form. If they qualify at this level, the next set of questions concerns a specific study that is appropriate for the presenting condition. The entire prescreening process takes about 15 minutes. An appointment is made for an intake interview scheduled within 24-48 hours.

having first access to subjects for his or her studies, and this may be a source of bias. A site should have a systematic means to rotate through candidates if multiple studies with similar criteria are ongoing. The screening procedure for inclusion and exclusion then becomes stepped, as follows.

1. Summary demographic information and statement of problem.

2. Designation for screening for a particular study or rotation for a class of studies.
3. Screening to determine appropriateness before making intake appointment.

4. Return to the general screening pool if the potential subject is found to be ineligible for the study for which he/she was first screened.

The extent to which preferential recruitment for one or another study introduces bias is unknown. However, it is intuitively sound to avoid the possibility of assignment bias whenever possible.

INTAKE

The intake procedures in all studies should be pedestrian but rigorous, systematic, and unbiased, protective of subjects rights, and informative of subjects' responsibilities. Failure in this domain can produce high dropout rates and produce results that cannot be replicated.

Consent Procedures

Sites differ in the characteristics of the consent procedures since variation within the broad National Institutes of Health (NIH) guidelines is permissible under the arrangement of local Initial Review Boards (IRBs). Requirements vary for style and other features of advertisements, consent forms, and supplementary information, despite the common required elements of consent procedures. The likelihood of this having an effect on recruitment, screening, and intake procedures is unknown.

Variation in initial intake may occur at a site due to differences across intake personnel and most certainly occur across sites. Sources of within-site differences may be due to different approaches to candidates based on biases involving perceived differences among patients. Thus, potential subjects may receive more or less information depending on unspecified ad libitum criteria imposed by the intake staff members (e.g., perceptions of intelligence, affluence, etc.). Thorough, well-documented consent forms, intake procedures, regular training and retraining, and relatively inflexible interviewing guidelines should minimize these problems. Further, monitoring for consistency within and across staff members should occur. This can be accomplished through regular audiotaping of consent, intake, and

other sessions. Regular meetings of intake and diagnostic staff members may help preclude drift in adherence to the criteria applied during the course of a study.

Since local IRB criteria may produce differences in length or other characteristics of consent forms, greater consistency can be achieved by having supplemental descriptive material, though this too must be submitted to the IRB. The material may be a useful additional guide to a patient after he or she has departed the premises. It should include information about fixed appointment days, times, and detail regarding provisions for continued participation as well as reimbursement. Here, as in other procedures, the goal is to reduce unnecessary variability in the experiences of subjects/patients entering a study.

When the agreement to participate has been obtained, the evaluative intake process is another source in which variability may arise. Consistency of measurement and application of diagnostic criteria is essential.

Diagnoses

Specific instruments used in diagnoses have been discussed at length elsewhere in other volumes in the NIDA Research Monograph Series. However, two points should be made. First, recall that many instruments were standardized on populations that may be rather different from the drug-dependent population. Further, some, such as the ASI, were developed using populations that may or may not be representative of the broader treatment-seeking population. Second, from the point of view of research and data analysis there should be an effort to minimize the number of instruments. This quest resides in the simple problem that increasing the number of measures and items increases the opportunity for statistically significant but spurious and clinically irrelevant findings. Assuming that appropriate instruments are being used, the critical issue is that inclusion or exclusion relies on the standardized criteria. The authors' intake procedures included:

A. Prescreening for general acceptability for study as described above.

B. Intake screen.

1. Complete medical evaluation including HIV, TB, EKG, drug screens, and other standard tests.
2. Complete psychiatric/behavioral evaluation including SCID, ASI, POMS, Hamilton A/D, and detailed drug history.

3. HIV testing/counseling-HIV risk behaviors.

4. Self-report instruments: POMS, Beck, desire to use drugs (craving).

If at any point during the intake process a potential subject is to be excluded from his or her assignment, there are two possible outcomes. First, the individual may be appropriate for evaluation for another ongoing study. At this point the intake continues, including any special items appropriate to the new assignment. If the subject cannot be included in any of the available studies, he or she has the opportunity to meet with a therapist to arrange for referral to other treatment sites.

A special problem emerges in the domain of substance use disorders. Oddly, because of the imprecise use of language, implementation of studies sometimes falls victim to words such as “abuse” when “dependence” is intended, and vice versa. Precision in terms of daily discussion among staff members should be encouraged since it reduces confusion and contributes to the integrity of the subject intake and baseline assessment procedures. Beyond this there is the lay vernacular, which permeates the field. In its 1994 publication guidelines, NIDA explicitly noted that terms with pejorative baggage, such as “addict,” should be avoided, and this applies as well to the extant clinic and staff meeting vocabularies. There are no parallels evident in other domains of medicine or psychology, yet the problem of applying nontechnical terms to drug-dependent patients is common in the professional community. Care in language may also contribute to better educating the subject about the disorder and thus real benefit can accrue to the patient as a research subject. Prior to each study, all intake staff members and clinicians as a group must review the conditions and criteria for entry again with a view to assuring that there is familiarity with the study-specific procedures (i.e., that they are implementing the same study).

Further, the authors have found it useful to be exclusive in screening with respect to diagnosis, preferring to err on the side of not including subjects for whom the diagnosis is less clear. Inclusion of dual-diagnosis patients can occur and have obvious undesirable consequences in a study intended to focus on one disorder, e.g., cocaine dependence. An extremely

heterogeneous patient population with a variety of secondary disorders, e.g., depression, antisocial personality disorder, and no ability or intent to stratify, may produce substantial variability. Clearly, this may be particularly problematic in an early efficacy trial, though it may be more acceptable in effectiveness studies, which typically account for this using a variety of procedures including larger sample sizes. As a matter of comparison to other areas, inclusion criteria in studies of antihypertensive medications, dermatological preparations, or other medications for medical conditions, tend to be characterized by considerably less variability for confounding conditions than is often found in studies of medications for drug dependence. Though it is often argued that users of single substances are rare, the authors have found persistence in recruitment can result in an adequate sample of subjects meeting the specified requirements and intent of the study without compromising the inclusion/exclusion criteria.

Heterogeneity of patient populations and differences across sites may have contributed to the equivocal results reported in the literature for some medications for drug dependence. For example, some studies reported modest or great improvement and still others report no change with desipramine (Arndt et al. 1992; Gawin and Kleber 1984; Gawin et al. 1989). Similar uncertainty has arisen in the case of fluoxetine (Batki et al. 1993; Grabowski et al. 1995). Characteristically excluding patients who have additional diagnoses (depression, antisocial personality disorder) or secondary conditions (AIDS) other than the specific drug dependence of interest may contribute to definitive findings in both efficacy and effectiveness trials (Grabowski et al. 1995). Arguably, costs are increased at the front end of a study due additional screening to achieve the desired sample. Nonetheless, it appears worthwhile to reduce variability to permit focus on the key issue; i.e., does medication X, under setting conditions A and B, and behavior therapy conditions C and D, produce benefit, no effect, or harm. Unambiguous criteria must be determined and applied during the initial screening, intake, and stabilization phases, and continuity must be sustained, often over a period of many months or several years.

Urine Drug Screens

Drug screens provide a critical element in defining the characteristics of the patient population at entry and during stabilization. Clinical and research staff must emphasize the importance of these data and the need for care in collection, transport, testing, and reporting (Hawks and Chiang 1987). Having a professionally constituted analytical chemistry laboratory on site provides greater assurance of reliability. Equally

important for behaviorally based studies is that an onsite facility provides for immediate results when called for by contingency management procedures. Not all sites can afford or require this level of participation by chemists and other technical personnel on site. Offsite laboratories providing slower turnaround times may be satisfactory during ongoing standard medication clinical trials. However, this necessarily slows the process of study entry, while awaiting the results of intake and stabilization drug screens. Slowing the intake process may in turn lead to failure of patients to return and slow overall study progress.

There is ongoing discussion of the type of screen required: qualitative, semiquantitative, or quantitative. In standard clinical trials semiquantitative urine screens should be sufficient and even qualitative results may suffice. Arguments for quantitative screens have emerged, but the supporting data for this position are not entirely persuasive. One concern is that qualitative or even semiquantitative screens must be interpreted in terms of cutoffs (e.g., 350 ng/mL). This was an arbitrary determination originally standardized for workplace screening where no use was acceptable and a minimal allowance for error was permitted. It is argued that a medication may reduce the level of dependence or abuse, but that this may go undetected unless quantitative screens are used. It could be argued that an effective medication (such as paralleling methadone in efficacy), would produce group reductions from 100 percent positive to 10 to 20 percent positive screens even by this stringent criteria. An alternative position is that a higher cutoff point could be used that would itself indicate relatively low levels of drug use. For example, since cocaine-dependent or -abusing patients often have benzoylecognine levels between 100,000 and 1 million ng/mL, a cutoff of 5,000 to 10,000 ng/mL would reflect significant change for most groups of patients.

The perspectives represented in the ongoing debate reflect conceptual as well as practical shifts in thought. There is increasing recognition that both risk reduction and risk elimination are important, with the former being satisfactory when the latter cannot be achieved. Whether strict elimination or risk reduction views are held, definitive and consistent criteria and effective procedures must be established and maintained at intake and stabilization as well as throughout the study. Drift in procedures can occur over the course of clinical trials that may take several years to complete. Permitting consistent comparison of data over time and within and across subjects must be avoided.
Other problems emerge with respect to medication-taking behavior and drug screens. To accommodate this, it may be useful and cost effective to differentiate phases of evaluation with respect to the comprehensiveness of screens. Given a relatively high level of tricyclic antidepressant use in the

community of cocaine-dependent patients (approximately 8 percent) the authors screened for both drugs of abuse and a full range of therapeutic medications in the initial screens. Reviewing this process, it was determined that an acceptable level of safety and rigor can be achieved by conducting comprehensive semiquantitative drug screens for common psychiatric medications at intake and monthly thereafter, while the twice-weekly (or more frequent) screens during study are restricted to commonly abused drugs.

In discussing the quantitative-qualitative issue, it may be necessary in early trials or certain types of combined medication behavioral therapy trials to obtain quantitative screens. However, semiquantitative screens should be adequate. The adequacy of this approach is testable. The authors' approach is that in a new series of studies quantitative screens are conducted for all drugs on all patients, and then varying cutoffs and criteria and results with respect to clinical utility are compared and applied. Future screening procedures will depend on the outcome of this comparison. Periodic blood screens may be useful for determining medication levels. At the same time, several studies in the literature, including the authors' work with fluoxetine, suggest little relationship between clinical effect and blood levels of commonly examined therapeutic medications. Clearly, when examining new medications, this issue must be evaluated.

Medical Evaluation

The medical evaluation while standard requires special attention to preclude unnecessary exclusion of potential subjects. Patients should generally be in good health except for problems directly related to drug use. Problems at intake may include results indicating aberrant EKG or liver dysfunction, placing a patient in a position of being borderline acceptable for a study. However, if the patient is otherwise acceptable, monitoring during the stabilization period will either mitigate concerns as symptoms abate or lead to exclusion. Dubious results that indicate greater than average risk as specified by the IRB human subjects provision require special attention. Thus, it may be necessary for a specialist, e.g., a cardiologist, to review a record before the patient may actually receive the study medication. Unlike many of the compliance issues, this matter is of greater concern with drug-dependent patients since they may take additional drugs while receiving the therapeutic agent. Even with an effective medication it can be expected that this will be particularly likely early in treatment, thereby increasing risk and potential harm. It is during baseline assessment that these issues must be addressed. Thus, in the authors' studies of stimulant replacement, EKGs were conducted three

times weekly in the first 2 weeks and once weekly thereafter. Sites unprepared, or unable to provide this level of evaluation, should be still more cautious in these initial evaluations.

Behavioral and Social Evaluation

Behavioral and social status/function evaluations are typically viewed as essential during the course of the baseline assessment. Data pertaining to these domains can be derived from standard diagnostic instruments, notably the ASI and SCID interviews. There appears to be increasing evidence that most demographic measures (e.g., race, income) have little relevance in examination of correlations with treatment outcome. Rather, factors proximate to drug use (e.g., drug, dose, severity, route), as well as comorbid psychiatric conditions may be most important. Certainly, a comprehensive drug history, particularly with respect to recent patterns of use, can be important and can be linked to the drug screen data. Further, the behavioral features of drug taking serve as the best dependent and independent variables. Of less clear value are measures such as dollars spent, grams used, and so on, unless they can be documented and validated against status of drug supplies in the community, (e.g., through the DEA). Drug prices and quality vary tremendously from time to time and across sources at the same time, thus diminishing even the face validity of such measures. While often reported, the generalizability or utility of such surrogate measures of drug use has yet to be demonstrated.

Useful data can be collected using queries focusing on patterns and circumstances of drug use. The Drug Use Desire Inventory, which relies extensively on operational definitions of behaviors thought to reflect craving, is used in the SARC Clinic. The principal problem with such measures is establishing definitive linkage to actual drug use. Particularly problematic, though conceptually interesting, is the not uncommon result of divergence of drug taking and self-report measures of desire to use as has been reported by Fischman and colleagues (1990), with desipramine.

Summary

The process of assuring quality and consistency in these phases of research is iterative and developmental; each member of the group contributes to inservice training on components of the process. Effectively, a checklist is developed assuring that all components are in place and agreed to before studies commence. The screening and intake process should be viewed as mechanical, with little room for error, producing data on which success of the results hinge. Each individual involved in the process can inadvertently tinker and contribute variability.

Many baseline assessment data can be directly entered into a computerized database, while others must be entered by hand at the earliest possible time. All files, papers, and computers should be sampled for accuracy. Further, all files must be retained to permit retrospective checks as needed. The authors have developed a computerized network system with terminals at the pharmacy medication-dispensing window, in intake interview offices, with research assistants, and in data coordinators' offices, to provide for regular patient checks and data entry. Messages to patients flow easily between clients and staff they have contact with to assure intake elements are completed and medicating sequences are initiated. Making a brief summary of the intake procedures produces an interactive process that permits improvements and minimizes errors. Beyond this a clinic operations manual should be available for all new staff members and should be reviewed periodically by all staff members to keep the manual procedures current.

STABILIZATION AND STUDY ENTRY

2 Weeks of Stabilization

Patients accepted for studies undergo a stabilization period prior to study entry. The scientific and practical advantages and disadvantages are discussed below.

As outlined earlier, there is concerted effort to assure that all patients receive the same information, agree to the same requirements of participation, and receive similar treatment at entry. Beyond this, however, there is a need to verify that the initial determinations are accurate. Acutely, it is to the investigator's disadvantage to need to reexamine since it is costly and may result in discharge of subjects; in the longer term it assures the validity of the sample and the results of costly and time-consuming clinical trials.

Evaluations During Stabilization

Patients should be monitored closely for a period of 10 to 14 days after intake. Study requirements and complexity of expected problems may result in clinic attendance from 2 to 7 days per week. Drug use may likewise be monitored through two to seven scheduled urine screens per week. Immediate return of urine screen results (within an hour) may be necessary in some cases, but return by the next visit is essential. Medication doses should be increased systematically during this period and the consequences of dosing observed; this will vary across

medications. For fluoxetine, ritanserin, risperidone, and methylphenidate studies, the authors considered it appropriate to obtain additional EKGs, while in methadone studies this requirement was not included.

Patients must be monitored to determine whether conditions apparent at entry such as depression wane during this period, and to determine whether previously unobserved symptoms emerge. It has been noted in the literature (e.g., Blaine et al. 1994; Kadden et al. 1995) that psychiatric diagnoses should be reassessed to determine whether an observed condition was stable or an artifact immediately preceding drug use. Ostensible coexisting depression is commonly noted in individuals who have recently ceased using cocaine, but in 60 percent or more of the cases, reevaluation demonstrates that the dual diagnosis disappears within 2 weeks. While demographic factors seem to have little bearing on outcome, comorbid psychiatric conditions or their absence does seem to be important with respect to treatment outcome, and thus with respect to the conduct and results of the clinical trial.

Finally, a fixed series of general queries should be posed at each session during this period concerning changes in legal status, living conditions, and social status in terms of significant others. These can warn of potential problems, assure that the patient continues to meet study criteria, and can also be checked against intake data derived from measures such as the ASI. Assuring that the subject clearly understands the requirements of the study can be accomplished by through repetition during the stabilization period through formal and informal means. It may be necessary to revisit consent procedures if it becomes apparent that the patient was not intact during the initial study introduction (Grabowski et al. 1979).

Therapists, research assistants, nurses, and any other staff members having contact with subjects can establish the framework for patients' participation during the stabilization period, making certain that fixed appointment conditions are met, that urine screens are delivered, and that materials are completed. Patients can be given a printed description that includes their regularly scheduled appointment time, other scheduled events such as urine screens, and delineation of items for which they receive research payments. These efforts promote a baseline level of compliance on which medications and behavioral therapy combinations can be evaluated. These constructive procedures are standard of care in some clinical settings but are rarely used in drug dependence research, where compliance is essential to rigorous evaluation.

Other Factors Confounding Assessment and Treatment Studies

Disregarding a long list of minor factors that may confound baseline assessment specifically and treatment research projects generally, there remain examples of major issues that may dramatically affect results. An illustrative example is provided here that is typically ignored, accepted with resignation, or encouraged and defended by many in the field of drug dependence. This is the issue of patient/subject attendance of self-help groups in the community and outside the control and purview of the study.

The baseline assessment is an important opportunity to determine whether other factors may confound the basic study of medication (or behavioral) therapy efficacy. The problem with self-help groups is clear. When evaluating a medication for hypertension, diet modification, exercise behavior, or seeking other treatments during the course of the study would be discouraged. Thus, researchers are particularly attentive to the issue of alternative ongoing therapies of any form for several reasons. Baseline data collection will be distorted in unknown ways by these activities. Beyond this, encouraging or not dissuading patients from extracurricular treatment activities assures a source of confound of unknown dimensions. During the stabilization period, researchers consistently emphasize the importance of adherence to the current therapeutic program. It is clear that the other activities may or may not be helpful but that the subject has agreed to participate in a specified treatment regimen for a defined period. In brief, researchers are committed to providing a particular range of treatment and committed to complying with this regimen. A variety of strategies are used. Remarkably, investigators conducting evaluations of medications for drug dependence both fail to discourage and may encourage attendance at self-help groups. Arguably, patients can seek to deceive the investigator. The authors view the effort at educating and obtaining compliance on this issue critical. Other sites prefer to account for this in other ways but only infrequently report the data. This single factor may contribute substantially to some problems observed in the literature. So-called self-help groups can have positive or negative effects and vary widely in focus, format, and extent to which they alter behavior. Often the message conveyed therein is directly contradictory to some cognitive behavioral strategies. The authors feel strongly that there is a need to assure that the treatment being evaluated is to the extent possible is the one being delivered at the study site. Again, by analogy, if patients in the hypertension study or a psychotherapy study were receiving prescription medications elsewhere or were self-medicating with active OTC medications, it would be cause for exclusion. The same should apply to supplemental doses of self-help groups. Other sources of variability are much more widely recognized and accounted for and will not be

addressed here. The example of self- help groups is emblematic of some of the problematic issues that confront the field and must be considered in the baseline assessment phase.

Another problem is that of accepting intent to treat as an essential criterion. At the extreme, it proposes that every subject who enters the clinic and signs a consent form must be included in subsequent analyses since there was a so-called intent to treat. The question is: Intent to treat what? Should misdiagnosed individuals be included? For example, in a study of depression+cocaine dependence should subjects be included whose depression lifts after several days? Or should patients who do not tolerate a dose of a widely used medication be considered as failures? The liabilities of this strategy are considerable. An obvious potential problem will come from rejecting medication or behavioral therapies that are effective.

Other problems common to the study of drug dependence treatment research, baseline evaluations, and design result from common myths or untested assumptions that are woven into the fabric of clinical trials. The problem of self-help groups has been mentioned; the view that patients must hit bottom is sacrosanct only in drug dependence and would be anathema in any other domain of medicine; the views regarding optimal setting conditions (e.g., inpatient, long duration therapy); that drug dependence is not fundamentally a real biological/psychological disorder: all contribute to confounds in efforts to develop optimal treatments. While there is an increasing body of literature contesting these beliefs, they present continuing challenges in the objective study of substance use disorders. These are apparent problems to the extent that these views permeate the views of staff members conducting the intake process.

Study Entry

The subject who completes the 2-week stabilization period with a stable diagnosis, with all other medical and psychiatric criteria met, and accepting all other conditions of study inclusion enters study and is not replaced. The importance of this continuing assessment period is exemplified in studies whether fixed or variable dosing prevails. The issue is particularly important in the former case such as the authors' fluoxetine study involving a placebo, 20 mg, and 40 mg of medication (Grabowski et al. 1995). If subjects were considered entered to a study before it is was determined that they can tolerate the assigned dose, differential dropout may skew the results. The same problem applies to other factors as well. For example, patients may state that they can attend a clinic 2 or 5 days per week but differential attrition may prevail for

working patients assigned to the condition, requiring more frequent visits. Again the results will be skewed. Many patients enter a study with liver function values that are borderline; the stabilization period permits determination of whether the values are stable, improving, or deteriorating (thus making the subject unacceptable for inclusion). In sum, the stabilization period permits evaluation of the practicability of study conditions for a particular patient while also providing for monitoring of the validity of the initial intake assessment.

When the subject is considered an active study participant a standard metric is applied to continued participation. The subject must sustain a level of 75 percent of his or her study commitments (e.g., urine screens, self-report form sessions, medication visits, therapy appointments). The actual percentage is arbitrary and should be established for the study and be somewhat flexible. In practice, the main function of this criterion is to provide a definitive endpoint for patients who drop out.

Problems

There are obvious problems with the stabilization strategy. For example, in the normal course of events, many subjects/patients leave treatment shortly after entry or within the first 2 weeks. Thus, apparent retention may be inflated if patients are not considered subjects until they have stabilized. This can be accommodated in the data-analytic process and description where the progression of attrition should be noted. For example, the number of individuals who underwent initial prescreening can be specified: those who entered intake but dropped out or were excluded and those who dropped out during stabilization. Data should be maintained for all patients and examined for differences among and between individuals who departed during stabilization and those who remained to become active subjects. Again, in a large study of fluoxetine, this procedure was found to be effective and there were no significant differences on the measures used between patients who departed during stabilization and those who were retained. Another obvious problem resides in added cost; however, it should be apparent that careful screening, albeit costly, is ultimately one of the most cost-effective features of the study process.

Summary

Despite the seeming complexity of the procedures described, they have proven generally acceptable in the authors' studies. Much of the mechanical character of the process is transparent to the patient. Precision is requisite for the difficult area of study comprised of clinical trials to

determine the efficacy of medications for drug dependence. There are considerations in medications development clinical trials that are overlooked. These include reevaluation postentry; prestudy to permit replicable comparison throughout the study, and finally assuring that the treatment being evaluated is the only one that the patient is receiving.

In the fluoxetine trial used as an example to this point, approximately 500 patients called (or walked in) and were thus screened using the telephone screening form. Most screened out at this level were polydrug users or had legal charges pending that might have interfered with participation. About 228 went through the intake procedure. Ultimately, 156 were completed stabilization and were formally considered having started study. The stringent requirements described to this point were applied. The end result was an uncompromised double-blind trial (Grabowski et al. 1995). Randomization has been successful in all of the authors' studies to date using these procedures. It is possible that consis-tently similar results could be obtained with less rigorous procedures; however, less rigorous or inconsistent intake procedures may well contribute to equivocal results in the literature.

OPTIMAL DESIGNS

Comment on design issues is warranted here at two levels, in addition to those by Nunes (this volume). First, it appears that despite flaws, the stabilization period as a formal study component is essential in this population, at the current level of understanding of the conduct of medications development trials. As noted earlier, this is particularly important to assure that all patients identified as such received the full dose. This was and is critical to evaluation of clinical efficacy and effectiveness.

Beyond this, it is suggested that variations in experimental designs should enhance detection of benefit, lack of change, or harm in medication trials. Medication trials for drug dependence are confronted with the issue of complexity of the disorder. Elaborate behavioral treatments as a baseline could conceivably obliterate differences between groups attributable to medications. Yet it is recognized that joint actions between behavioral therapies and medication may enhance effectiveness. These can be approached as two distinct types of studies. They may at times be examined concurrently, in the following manner: An extended (e.g., 6-week) double- blind baseline period with placebo and medication using standard care (e.g., one therapy session per week) could provide a rigorous test of the medication. Substantive reinforcers could be provided for retention but not contingent on reductions in drug use. If an effect is

observed under the austere standard condition, the medication might be viewed as an important candidate for further examination. During such a period, a medication such as methadone would readily be determined to reduce opiate use. If no difference was observed during this period, the medication might be sacrificed as a candidate. If pronounced or even modest differences are observed, application of intensive behavioral interventions could be applied to half of the initial remaining subjects in each group. This would permit evaluation of the extent to which a behavior therapy medication interaction produces further change. Disproportions in group size could emerge in the second phase of this design if the medication was effective. However, this design or some other hybrid could greatly reduce the cost, time, and steps involved in initial trials of efficacy and effectiveness.

OPTIMAL DATA-ANALYTIC STRATEGIES

The authors are currently examining optimal data-analytic strategies permitting capture of the best possible baseline assessment data for comparison to later progress through the study. New analytic tools are being evaluated for consideration of dropouts, missing data, and other hazards of this research. By maintaining records at each stage of the process—from initial screening onward—comparisons are feasible.

Considerable concern has emerged regarding the adequacy of commonly used measures. Psychometrics must be impeccable for obvious reasons but problems do emerge. For example, the authors have found that factor analytic strategies with the POMS may create problems in this field since the factors wash out on careful scrutiny. This may be due to the population on which it was standardized and the comparisons that are being made. In one analysis, using education as a surrogate variable for reading ability, the authors found that the factors can be isolated for those who have a 12th-grade reading level or higher, but not if less education than 12th grade. This suggests that there may be further problems with other measures adapted from other psychiatric populations. Thus, these measures may be inappropriate to detect the changes at a later time.

Beyond this, the utility of using many surrogate measures and attempting to identify predictors must be considered. As previously noted certain key variables such as severity appear to be important while many demographic variables are of limited or no value. At this point, the field would do well to focus on the main task of developing effective treatments for the substance use disorder. While there is rarely such correlation seeking in other areas of medicine, it does emerge in other areas of psychological disorders such as panic and phobias. In these fields there have been calls

for a return to the focus on the core disorder, and the advice would seem to apply to substance use disorders as well.

SUMMARY AND CONCLUSIONS

To summarize, perhaps most important, but most difficult to achieve will be commonalties and standardization across trials so that rigorous comparison is possible. Researchers will do well to examine clinical trials in other areas, mimicking those elements that are compatible, avoiding those that are not, and above all, avoiding costly reinvention.

REFERENCES

Arndt, I.O.; Dorozynsky, L.; Woody, E.; McLellan, A.T.; and O'Brien, C.P. Desipramine treatment of cocaine dependence in methadone maintained patients. *Arch Gen Psychiatry* 49:888-893, 1992.

Batki, S.L.; Manfredi, L.B.; Jacob, P.; and Jones, R.T. Fluoxetine for cocaine dependence in methadone maintenance: Quantitative plasma and urine cocaine/benzoylecognine concentrations. *J Clin Psychopharmacol* 13:243-250, 1993.

Blaine, J.; Ling, W.; Kosten, T.R.; O'Brien, C.P.; and Chiarello, R. Establishing the efficacy and safety of medications for the treatment of drug dependence and abuse: Methodologic issues. In: Prien, R., and Robinson, D., eds. *Clinical Evaluation of Psychotropic Drugs: Principles and Guidelines.* New York: Raven Press, 1994.

Elk, R.; Grabowski, J.; Rhoades, H.; and McLellan, A.T. A substance abuse research-treatment clinic. *Subst Abuse Treat* 10:459-471, 1993.

Fischman, M.W.; Foltin, R.W.; Nestadt, G.; and Pearlson, G.D. Effects of desipramine maintenance on cocaine self-administration by humans. *J Pharmacol Exp Ther* 253:760-770, 1990.

Gawin, F.H., and Kleber, H.D. Cocaine abuse treatment-open pilot trial with desipramine and lithium carbonate. *Arch Gen Psychiatry* 41:903-909, 1984.

Gawin, F.H.; Kleber, H.D.; Byck, R.; Rounsaville, B.J.; Kosten, T.R.; Jatlow, P.I.; and Morgan, C. Desipramine facilitation of initial cocaine abstinence. *Arch Gen Psychiatry* 46:117-121, 1989.

Grabowski, J.; Rhoades, H.; Elk, R.; Schmitz, J.; and Creson, D. Clinic-wide and individualized behavioral interventions in drug dependence treatment. In: Onken, L.; Blaine, J.; and

Boren, J., eds. *Behavioral Treatments for Drug Abuse and Dependence.* National Institute on Drug Abuse Research Monograph 137. NIH Pub. No. 93-3684. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1993. pp. 37-72.

Grabowski, J.; Rhoades, H.; Elk, R.; Schmitz, J.M.; Davis, C.; Creson, D.; and Kirby, K. Fluoxetine is ineffective for treatment of cocaine dependence or concurrent opiate and cocaine dependence: Two placebo-controlled, double-blind trials. *J Clin Psychopharmacol* 15:163-174, 1995.

Grabowski, J.; O'Brien, C.P.; Greenstein, R.; Long, M.; Steinberg-Donato, S.; and Ternes, J. Effects of contingent payment on a naltrexone regimen. *Am J Drug Alcohol Abuse* 6:355-365, 1979.

Hawks, R., and Chiang, C.N., eds. *Urine Testing for Drugs of Abuse.* National Institute on Drug Abuse Research Monograph 73. DHHS Pub. No. (ADM)87-1481. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1987.

Kadden, R.M.; Kranzier, H.R.; and Rounsaville, B.J. Validity of distinction between substance induced and independent depression and anxiety disorders. *Am J Addict* 4:107-117, 1995.

Kartzinel, R.; Lisook, R.; Rullo, B.; Severe, J.; and Spilker, B. Clinical trial implementation. In: Prien, R., and Robinson, D., eds. *Clinical Evaluation of Psychotropic Drugs: Principles and Guidelines.* New York: Raven Press, 1994. pp. 161-184.

Kupfer, D.; Kraemer, H.; and Bartko, J. Documenting and reporting the study results of a randomized clinical trial (spicy meatballs not pabulum). In: Prien, R., and Robinson, D., eds. *Clinical Evaluation of Psychotropic Drugs: Principles and Guidelines.* New York: Raven Press, 1994. pp. 237-260.

Levison, P.; Gerstein, D.; and Maloff, D. *Commonalities in Substance Abuse and Habitual Behavior*. Lexington, MA: Lexington Books, 1983.

O'Brien, C.P., and McLellan, A.T. Myths about the treatment of addiction. *Lancet* 347:237-240, 1996.

Prien, R., and Robinson, D., eds. *Clinical Evaluation of Psychotropic Drugs: Principles and Guidelines.* New York: Raven Press, 1994.

AUTHOR

John Grabowski, Ph.D.
Director, Substance Abuse Research Center
Professor, Psychiatry and Behavioral Sciences
Department of Psychiatry
University of Texas
1300 Moursund Street
Houston, TX 77030

Click here to go to page 182