

**CONSENSUS STATEMENT ON EVALUATION OF OUTCOME OF
PHARMACOTHERAPY FOR SUBSTANCE ABUSE/DEPENDENCE**

**REPORT FROM A NIDA/CPDD MEETING
held in Washington, DC**

April 23-24, 1999

Meeting organizers: Frank Vocci, NIDA Medications Development Division
Harriet de Wit, College on Problems of Drug Dependence

Participants listed in appendix.

GOALS

This meeting was held to accomplish two main goals: i) to achieve a consensus about the critical features of pharmacotherapy trials to be used in the treatment of nicotine, alcohol and stimulant abuse/ dependence, and ii) to write a document outlining these features to guide in the design and evaluation of future pharmacotherapy trials.

The participants addressed the following key questions:

1. What should be the primary outcome measures in a pharmacotherapy trial for nicotine, alcohol or cocaine abuse? What secondary measures should be obtained, and how should these be used in light of the primary measures?
2. What are realistic treatment goals for smoking cessation, alcohol treatment or treatment of cocaine abuse? What weight should be given to reduction in use, as distinct from abstinence? How can use or lapse be distinguished from relapse in a clinical trial?
3. Should every pharmacotherapy trial include a psychosocial component? If so, how can this be standardized?
4. What are the most appropriate statistics to evaluate outcome, and how should dropouts be handled?

Participants divided into four discussion groups. Three groups discussed outcome measures related to treatment of alcohol, cocaine and cigarette dependence, respectively, and a fourth group discussed statistical issues related to assessment of treatment outcome. The following are the reports of the four groups' discussions.

1. OUTCOME MEASURES FOR SMOKING CESSATION

Chair: John Hughes

Rapporteur: Ken Perkins

Medication trials in the smoking cessation area have been more numerous compared to alcohol and cocaine. Thus, the focus of our discussion was primarily on evaluation of the utility of commonly used goals and measures rather than identification of new outcomes or measures. This group rated their confidence in their recommendations, based on the available empirical evidence, in the following way: A confidence level of “A” indicates that a high level of confidence and strong empirical support, a confidence level of “B” indicates a moderate level of confidence, and a confidence level of “C” indicates that the group considered this recommendation to be helpful but lacking in empirical support.

a. Primary treatment goals and measures

In contrast with alcohol and cocaine, continuous abstinence is the most important general outcome in smoking cessation trials. Whether or not the definition of continuous abstinence should always require absolute abstinence (“not even a puff”) or allow for “slips” (intermittent smoking short of National Heart Blood and Lung Institute-defined relapse, which is 7 consecutive days of smoking) is not clear and requires continued research. This uncertainty revolves around lack of knowledge as to the sensitivity and specificity of lapses, particularly soon after the quit date, as a predictor of full-blown relapse. However, absolute abstinence with no slips is a reasonable requirement over the last 4 weeks of medication use in order to demonstrate initial efficacy or subsequent maintenance of efficacy/relapse prevention.

Five specific primary treatment goals were identified:

- 1) initial cessation: continuous abstinence (no slips) over the last 4 weeks of medication treatment as of an a priori determined point in follow-up (e.g. 6 or 8 weeks after initiation of medication, prior to any tapering phase). Confidence in the appropriateness of this as an outcome goal and in its definition--A.
- 2) maintenance of efficacy: continuous abstinence over at least 6 months' duration of medication use, including the last 4 weeks with no slips, among those achieving initial cessation (see #1 above) using the medication of interest. Confidence in appropriateness as an outcome goal--A, in the definition--B.
- 3) relapse prevention: continuous abstinence over at least 6 months' duration of medication use, including the last 4 weeks with no slips, among those achieving initial cessation using any treatment. If initial cessation is achieved with the medication of interest, a drug-free interval of some period (e.g. 1 month) should intervene before resuming the medication in order to determine relapse prevention. (Otherwise, the trial would be a test of maintenance of efficacy.) Confidence in appropriateness as an outcome goal--C, because of virtual absence of research on medications for this indication; confidence in the definition--B.
- 4) craving/withdrawal: Withdrawal is clearly defined by measures assessing DSM-IV withdrawal criteria, but craving is not clearly defined. The duration and magnitude of decrease in craving or withdrawal necessary to demonstrate a clinically significant change are not known.

There is uncertainty about the importance of these changes as a primary indication of a medication. Confidence in craving and withdrawal as primary outcome goals--B; confidence in measure of withdrawal--B or craving--C.

5) smoking reduction: Reduction in smoking is unclear as a primary outcome measure. While smoking reduction may eventually be a worthwhile and well-justified goal, there is currently insufficient research supporting its importance due to lack of data regarding the possible health benefits, difficulty in measuring reduction (e.g. lack of adequate biochemical verification in nicotine replacement trial), and whether smokers can maintain reduced smoking exposure over a long time period. Another potential problem is whether reduction as a treatment option may lead to fewer quit attempts among those who otherwise would attempt to quit completely. Confidence in smoking reduction as a primary outcome goal--C; confidence in measures when Nicotine Replacement Treatment (NRT) used--marginal; confidence in measures when treatment other than NRT used: C.

b. Secondary treatment goal

One secondary treatment goal was identified: craving/withdrawal. Definition and measures are discussed in #4 above. Confidence in each as a secondary outcome goal--A; confidence in measure of withdrawal--B of craving--C.

c. Biochemical verification

There is currently no adequate biochemical verification of long-term (one week or more) continuous abstinence. However, all smoking cessation trials should include some biochemical measure of abstinence in order to accurately gauge recent abstinence rates and because presence of such a measure is believed to increase the truthfulness of subject self-report of smoking. Cotinine assesses smoking exposure over several days, as opposed to the briefer period assessed by expired-air carbon monoxide (CO, 24 hrs or less). However, cotinine is generally not valid in trials of nicotine replacement. Thus, to equate measures across various medication trials, CO is thought the most appropriate measure. On the other hand, CO may be insensitive for detecting continued smoking in children and others with low smoking rates, and cotinine may be preferable in studies of those populations not involving nicotine replacement. Confidence in obtaining biochemical validation in all smoking cessation trials--A; confidence in measures--B.

Biochemical measures of smoking reduction are not yet available, although reduction in cotinine may be acceptable in non-nicotine replacement trials of smoking reduction. Confidence in biochemical measures for reduction--untested.

d. Psychosocial treatment

A wide range of psychosocial treatments to accompany medications may be appropriate, from very intensive to minimal and perhaps none. However, the type and duration of counseling should be commensurate with that likely to be provided in the setting in which the medication is intended to be used. For example, if the medication is intended for use in a primary care setting, psychosocial treatment should be comparable to what would likely be provided by a primary care office. Confidence in recommending a variety of levels of psychosocial treatments--A.

e. Missing data

Optimal handling of missing data continues to be unclear. Missing data should not automatically be viewed as indicative of relapse. An a priori algorithm to determine how missing data should be categorized (relapse, continued abstinence) based on subsequent information may improve accuracy of obtained abstinence rates. For example, it may be appropriate to code as continuously abstinent subjects who otherwise demonstrate continued abstinence before and after a missing appointment and report a reasonable excuse (e.g., notify the program in advance that they will be absent or on a business trip). However, research is needed to determine the validity of such an algorithm approach to handling missing data. Confidence in how to handle missing data--C.

f. Statistics

The primary measure for initial cessation, maintenance of abstinence, and relapse prevention is four weeks of continuous abstinence (no slips, not even a puff) at a predetermined assessment point post-quit date and no relapse (7 consecutive days of smoking) at any time during medication use. Appropriate secondary analyses are point prevalence (1 week of continuous abstinence even if prior relapse occurred) and continuous abstinence since initial quit (see outcome #2 above), which are both well-accepted dichotomous outcomes appropriate for non-parametric statistics. Survival analysis of days to relapse (7 consecutive days of any smoking) or first lapse (any smoking) following initial quit date can provide useful information on duration of treatment efficacy and is also appropriate as a secondary outcome measure. Confidence in 4 weeks of continuous abstinence as primary outcome--A. Among secondary outcomes, confidence in point prevalence and continuous abstinence--A; confidence in survival analysis--B.

2. OUTCOME MEASURES FOR ALCOHOL ABUSE AND DEPENDENCE

Chair: Damaris Rohsenow

Rapporteur: Stephanie O Malley

There is a recognized need for a uniform set of outcome criteria in the treatment of alcohol abuse and dependence. The Institute of Medicine (1990) report on treatments for alcohol problems did not specify a set of outcome criteria that should be used, but it was recognized that outcome measures should reflect diverse aspects of functioning, including but not limited to levels of drinking. The report recognized that many studies have relied primarily on changes in the level of alcohol use to determine the value of a particular form of treatment and could benefit from other outcome measures. The report recommended that “a uniform set of outcome criteria would be an enormous advantage” (p. 321), to allow comparisons of treatments across different studies.

a. Primary outcome measures

A variety of outcome criteria were discussed by the alcohol subgroup. The consensus was that the best outcomes to assess for treatment trials are the following continuous variables. It was assumed that patients would initially be abstinent at the start of treatment as a condition for participation.

i. *Percent heavy drinking days* A heavy drinking day is defined generally as any day in which a person consumes more than a certain number of standard drinks. (The size of standard drinks is defined as by the criteria in the Time Line Follow Back interview, below.) In general, the minimum criteria accepted for a heavy drinking day are either 5 or 6 drinks on a day for men and either 4 or 5 drinks per day for women. The percent of heavy drinking days is the number of days that meet these criteria divided by the number of days in the follow-up period, excluding days on which the subject was incarcerated (Sobell, Sobell & Ward, 1980; referred to as the number of possible drinking days).

The advantage of percent heavy drinking days as a primary outcome variable is that it incorporates abstinence as well as severity in a single metric. That is, people who are completely abstinent for the entire period will score zero percent heavy drinking days, and people who have many completely abstinent days will score low on this variable. In addition, people who are drinking lightly will also score low on this variable. Thus, a low value on this variable is a function of either frequent abstinence and/or infrequent heavy use. (Confidence: A.)

ii. *Percent days abstinent*. This variable is the number of abstinent days divided by either the number of days in the period or the number of possible drinking days. The advantage of this variable is that it includes people who have been completely abstinent throughout as well as people with varying amounts of abstinence in a way that reflects the continuum of frequency of drinking. One limitation of this variable is that its distribution is usually significantly skewed so that this variable needs to be transformed prior to analysis. (Confidence: A.)

iii. *Number of drinks per drinking day.* This variable is the total number of standard drinks consumed during the follow-up period divided by the number of days on which drinking occurred during the period. It is not clear how best to handle data from those people who have no drinking days, to avoid dividing by zero. Some researchers code these days as zeroes, and other researchers code these as missing data. The combination of percent days abstinent and number of drinks per drinking day (when no drinking is coded as missing) in a single study is beneficial because these two variables have a lower correlation with each other than many other possible pairs of variables do, and thus represent relatively independent dimensions of outcome. (Confidence: A.)

iv. *Days to first heavy drinking day.* This variable is the length of time between treatment and the first occurrence of a heavy drinking day, defined as above. The length of time that a patient does well after treatment before experiencing a relapse to heavy drinking is a variable of interest that is relatively independent of the other variables. This variable can be used in survival analyses to determine between-groups differences in survival functions. The advantage of survival analysis with this variable is that it includes both patients who did and did not have a heavy drinking day along with the length of time before heavy drinking occurred. This approach is particularly useful when there are high rates of loss to follow up or missing data after the first heavy drinking day has occurred. (Confidence: A.)

b. Secondary Outcome Measures

i. *Biochemical markers.* The best of the biochemical measures (GGT and CDT) provide reasonable concordance with heavy drinking as distinct from moderate drinking or abstinence over an extended period of time, at least in males. However, these indices provide essentially a dichotomous measure of relapse, and thus a less sensitive measure of change. Establishing a baseline for the biochemical marker and subsequently measuring the change in level of the marker during follow-up assessment may improve detection of heavy drinking.

ii. *Alcohol problems.* It is sometimes useful to know whether the patient is experiencing fewer or less severe alcohol-related problems as a result of a treatment. Reduction in alcohol-related problem severity is presumed to result from reduction or cessation of alcohol use rather than being a direct effect of pharmacotherapy independent of change in alcohol use. However: (1) Reductions in alcohol-related problems may lag behind reductions in alcohol use by 3 months or more, so this may be better as an indicator of longer-term outcome. (2) Reductions in alcohol-related problems may be relatively independent of reductions in alcohol use as a person may choose to drink in less risky situations without reducing use overall, or the person may cease alcohol use but have chronic or permanent problems that resulted from the use.

Currently, probably the best measure for the purpose is the Drinker Related Inventory of Consequences (DrInC-2L) from Project MATCH (Miller et al., 1995), consisting of 50 yes/no items that all fall on one reliable factor. The items can also be scored for five content areas: Physical, Social Responsibility, Impulse Control, Interpersonal, and Intrapersonal. (Confidence: B.)

iii. *Dichotomous measures of abstinence.* Another less sensitive outcome measure of interest to many people is a dichotomous measure of abstinence versus any use (“lapse”) or of

relapse versus any outcome other than a relapse. (1) Abstinence versus lapse is generally defined by scoring a patient who had even a single drink during the outcome period as having lapsed. (2) Relapse has been defined in various ways. A recommended way is to define a patient as relapsed if the patient had even a single heavy drinking day (as defined above). (Confidence: A.) Another way is to define relapse as the occurrence of two or three heavy drinking days during the outcome period. (Confidence: B.)

c. Measurement issues

i. Measures of hypothesized mediating mechanisms of effect. Different medications may have their effects by different hypothesized mechanisms, affecting different phases of the dependence process (e.g., reducing direct drug effects, withdrawal symptomatology or craving after abstinence). Outcome measures should be used that are appropriate to the purported effect of the medication. There is no agreement currently about the most effective way to do this, but several options have been suggested. For example, medications that are presumed to affect desire for alcohol by affecting general appetite or protracted withdrawal should include measures of these variables. Some medications may act by reducing the acute effects of alcohol and could be tested in a challenge paradigm whereas other drugs, such as disulfiram, would not be appropriate in an alcohol challenge study. (Confidence: B)

ii. Preferred assessment method. The measure that is currently the best means of assessing alcohol use treatment outcome is the Time Line Follow Back interview (Sobell, Sobell, & Ward, 1980). This is a calendar-assisted method of collecting self-report data back over periods of up to a year, although many investigators prefer not to exceed 6 months of retrospective data collection. The other preferred method is the Form 90 interview (Miller, 1996). These assessments should be conducted following certain guidelines (Sobell and Sobell, 1986) for maximizing the validity of self-report: These include creating a setting of acceptance and confidentiality, having the interviewer be different from the therapist, providing an expectation that there will be no disapproval of reports of drinking, assuring a zero blood alcohol level at the time of the interview, and collecting corroborating reports from a significant other (which can function as a bogus pipeline to increase validity, but may not be used to change the self-report data from the patient). (Confidence: A.)

iii. Need for a baseline. A baseline measure of pretreatment values of the drinking variables is necessary to characterize and control for individual differences in pretreatment functioning. Evidence indicates that the one month prior to entry into treatment is atypical of the levels of drinking that occur during the rest of the year before starting treatment, sometimes showing large increases that lead up to the decision to enter treatment, and sometimes showing large decreases as the person tries to cut down or stop on their own due to the motivation to change (Cooper, Sobell, Maisto, & Sobell, 1990). For this reason, a minimum of a 90-day or 180-day pretreatment baseline is recommended. (Confidence: A.)

iv. Other variables discussed. Some other variables were discussed but were not deemed to be as valuable as the primary measures of the effects of pharmacotherapies. (1) Dichotomous measures (such as relapsed versus abstinent) were not recommended to be primary as they have very low power to detect differences (Kraemer, in press) and as they combine

people with a single episode with people who have used continuously throughout without distinction and are thus relatively insensitive to change. (2) Global clinician assessment ratings were also not considered to be good primary measures. It is important that outcome measures be gathered by people who do not have a vested interest in the treatment to minimize a social desirability bias that occurs when a clinician reacts to the patient's reports with approval or suggestions for change (Institute of Medicine, 1990; Sobell, Sobell, & Ward, 1980). Because of the inherent bias, clinician ratings are less likely to be valid indicators of outcome than are self-reports of drinking when gathered under the recommended conditions. (3) Biochemical markers of drinking outcome are not yet recommended as the primary outcome measures because of their low sensitivity to change.

d. Realistic Treatment Goals

i. *Goals according to whom?* The goals of treatment can vary somewhat depending on the person or institution that is involved in setting goals. There are similarities and differences in treatment goals among alcohol treatment centers, the institutions that pay for treatment for alcoholics (commonly managed care companies and state governments), the alcoholics themselves, family members and employers of alcoholics, medical staff and police who deal with crashes and injuries, and society and government.

ii. *What are the goals?* The Institute on Medicine (1990; p. 313) report provides a useful conceptual framework for treatment goals, suggesting that a distinction can be made between the short-term goals of treatment and later outcome. The report goes on to say that some of the goals of treatment include detoxification (if needed); the reduction or elimination of alcohol ingestion; reduction in signs, symptoms, and consequences of alcohol use; reduction of concurrent medical, psychiatric, and social problems, and a changed attitude toward drinking in favor of a commitment to healthy change in drinking habits. Three phases of outcome from treatment can be distinguished: (1) detoxification, (2) active treatment to produce initial total abstinence, and (3) relapse prevention approaches to prevent return to problematic levels of drinking (Institute of Medicine, 1990, pp. 511-512).

iii. *Complete abstinence, permanently.* The primary goal of most treatment centers in the United States is complete abstinence without resumption of any drinking. Complete abstinence is a place to start alcoholics in treatment. However, it may not be a realistic treatment outcome at this time given the high rates of lapses and relapses during the first 6 months following any one treatment attempt.

iv. *Eliminate alcohol-related problems.* Alcoholics enter treatment because of problematic consequences that resulted from drinking. Without any problems, they would not want to change their drinking. Therefore, the ultimate treatment goal of alcoholics is to eliminate the problems. For many alcoholics, abstinence is the only means to this end. However, for other alcoholics, reductions in drinking are a means toward this end. Family members would not be likely to be concerned about the patient's drinking except for the problems the drinking resulted in, problems such as frequent intoxication, fights, arguments,

violence, and not handling responsibilities. Employers are concerned about problems on the job resulting from drinking, including lost productivity (missed days and slowdowns), poor quality work, and accidents on the job resulting in injuries to people or damage to equipment. Police are concerned about crashes, injuries, disruptive behavior, and crime from alcohol use. Society and government are concerned about health costs from illnesses and accidents, lost productivity in the workplace, crime, and the effects of alcohol on the family.

v. Improve psychosocial functioning. Abstinence without improved functioning in life is considered by some to be only half the battle won. Many alcoholics have concomitant impairments in work, family, health, and psychological functioning. Improvements in these areas are desired by alcoholics, family, employers, friends, and society.

vi. Reduce future treatment costs. Those who pay for substance abuse treatment - insurance providers, managed care firms, and state governments, primarily - have treatment cost reductions as their bottom line. This means their goal for substance abuse treatment for alcoholics is to eliminate future need for substance abuse treatment and to reduce future medical costs in general. Short term costs, i.e., treatment utilization within one or two years, are generally of primary concern to insurance payors rather than long term utilization over the patient's lifetime.

vii. Means to these ends. For many, complete abstinence is the only means to all of the ends listed above. However, it is possible that some amount of slips, light drinking, or infrequent drinking can occur while still allowing a reduction or cessation of alcohol-related problems, reduction of future treatment costs, and improved psychosocial functioning. For these reasons and because few alcoholics obtain permanent complete abstinence, reduction in quantity and frequency of alcohol use in general and in heavy drinking specifically is recommended as the most realistic treatment outcome to assess. (Confidence: A.)

e. Including a Psychosocial Treatment Component

It is recommended that every pharmacotherapy trial for alcohol dependence include some form of psychosocial treatment for all participants. (Confidence: A.) One reason for this is ethical, in that alcohol dependent patients in the placebo arm of the trial should be receiving some acceptable form of treatment. Even in an open-label trial, if the medication does not have an immediate beneficial effect for all patients, the patients need an active treatment to support them. A second reason is scientific, in that psychosocial treatments can assist in increasing compliance with medication, increasing retention in treatment, and be designed to help the patients deal with the many psychosocial sequelae of their alcoholism. Furthermore, providing a psychosocial treatment increases the generalizability of the results as the study can recruit a broader range of patients who are more representative of those in treatment programs. Without psychosocial treatment, a study might need to be considerably more selective about which patients may ethically be enrolled.

No one standard type of psychosocial treatment needs to be used across different pharmacotherapy trials. However, the nature of the psychosocial treatment needs to be

standardized within a pharmacotherapy trial. Standardization needs to be done to decrease “noise”: variance in outcomes attributable to differences in the way the psychosocial treatment is conducted. (Confidence: A.) Ideally, a treatment manual should be used specifying guidelines for the approach to treatment, but it does not need to be elaborate. Many adequate treatment manuals exist that could be adopted. (Confidence: A.) Training of therapists needs to be done if the study is conducted across sites or with more than one therapist administering the psychosocial intervention. (Confidence: A.) Monitoring of how the therapists administer the treatment should be done if the study is conducted across sites or with more than one therapist administering the psychosocial intervention. However, the system for monitoring does not need to be elaborate or expensive, as a checklist format asking therapists to rate their delivery of the treatment components may be acceptable. (Confidence: B for method.)

Several criteria should be used when choosing and delivering the psychosocial component. First, the psychosocial treatment needs to be one that is considered at least the minimally acceptable treatment for patients in the placebo arm of the study. Second, anyone who cannot tolerate the medication or who chooses not to continue with the medication for any reason should continue in the psychosocial treatment rather than being dropped from that treatment as well. For this reason, the treatment should be more than simply medication management so that patients can continue to receive an active treatment. The treatment should include at least educational advice about drinking. Third, it would be beneficial to the pharmacotherapy trial if the psychosocial treatment included discussion of medication compliance issues and management of side effects. It is recommended therefore that the treatment manual includes at minimum: (1) compliance issues, (2) side effects management and advice, (3) advice about avoiding drinking and maintaining sobriety. (Confidence: A.)

f. Statistical Issues about Dropouts and Analyzing Outcome

Using intent-to-treat analyses is the best approach and has become a standard in the treatment field. It is important to analyze outcomes for every patient who entered the medication trial, regardless of how rapidly dropout occurred. This means that it is crucial to collect information about the outcome variables throughout the data collection period even among those who dropped out of the medication trial. This can be done with a high rate of successful follow up if procedures are set in place from the start of the study to follow everyone possible. This includes explaining to patients up front the importance of completing the follow-up interview even if they discontinue treatment, obtaining contact information from the patients on two or more locaters as well as on the person to be used for collateral confirmation of self-report data, having the patient sign a letter to each of the locaters and collateral informant at the time of recruitment that can be used later, obtaining drivers’ license numbers or social security numbers of the patients, and obtaining any additional information such as work setting or halfway house address that may aid in tracking the patient. Payments for follow-up interviews need to be high enough to be a reasonable incentive for participation without being coercive. (Confidence: A.)

Missing data can occur for a variety of reasons. Patients can refuse to return for interviews because they are doing well and want to forget about their past drinking, or because

they are doing poorly. Therefore, assuming everyone lost to follow up is relapsed may be an overestimate of relapse rates although it is one choice to provide a conservative estimate of success. Continuous variables that are missing can be imputed statistically based on the data of the patients whose data are available using regression estimation procedures or growth models, if the statistical assumptions of the procedure are met. (Confidence: B.) However, any procedure for estimating missing data still requires that the amount of missing data be minimized, so it is highly important to obtain data from as many patients as possible.

It would be useful to conduct analyses that take into account the time from the initiation of treatment in a medication trial. Different medications may be particularly effective either early in treatment, or later. For some drugs, data from the first week or two of the medication trial might be excluded from the analyses. During this early time period, some patients may be testing the treatment, having brief early relapses or lapses, or be so motivated to change that no between-groups differences can appear. Nevertheless, it would be important to collect data throughout all phases of the trial. This approach of examining only certain phases of the time course of treatment is new and speculative at this time. (Confidence: C.)

Another way to consider the time course is to analyze data from different time periods rather than aggregating all data across the entire outcome collection period. Growth models are one method for doing this. Another is to analyze blocks of time either with separate analyses of covariance or logistic regressions, or entering the time factor as a variable in a mixed between-group and repeated measures analysis. (Confidence: B.)

The baseline data should be used as a covariate rather than as the basis for a change score for statistical reasons. First, change scores may have highly variable distributions that violate distributional assumptions of analyses. Second, change scores double the error variance in analyses, decreasing the likelihood of detecting real differences. Third, the ultimate question in treatment trials is a between-groups question rather than a within-groups or within-individual questions, and the between-groups comparisons can control for baseline differences using covariance methods (provided that the statistical assumptions are met). (Confidence: A.)

Some people have proposed a 50% reduction in drinking days or amount consumed per occasion from baseline to follow up as a criterion for success. This is not recommended for several reasons. First, this criterion requires change scores which are problematic as described above. Second, it converts a powerful continuous measure into a dichotomous outcome criterion, vastly reducing the power to detect differences. Third, it is highly dependent on baseline values, making the data very sample bound and decreasing generalizability. Fourth, because it is highly dependent on baseline values, it may not provide a valid indicator of whether an alcoholic is doing well: An alcoholic who had been drinking infrequently before treatment while struggling to become sober may be seen as a failure after treatment when drinking on very few days, while an alcoholic who had been drinking every day prior to treatment may be seen as a success although drinking on an unacceptably high number of days during the outcome period. Instead, a criterion that is consistent across patients within the outcome period and not dependent on baseline values is preferable, such as having not more than two or three heavy drinking days during any three-month window. The idea that a criterion

such as a 50% reduction in some measure of drinking should be a standard for accepting the efficacy of a pharmacotherapy needs further evaluation and discussion.

g. References

Cooper, A. M., Sobell, M. B., Maisto, S. A., & Sobell, L. C. (1980). Criterion intervals for pretreatment drinking measures in treatment evaluation. Journal of Studies on Alcohol, *41*, 1186-1195.

Institute of Medicine. (1990) Broadening the Base of Treatment for Alcohol Problems. Washington, DC: National Academy Press.

Kraemer, H. C. (in press). You can't fix by analysis what you bungle by design. In F. M. Tims, J. D. Blaine, L. S. Onken, & B. Tai (Eds.), Treatment of cocaine abuse: Outcome research. Monograph, National Institute on Drug Abuse.

Miller, W. R. (1996). FORM 90: test manual. Project MATCH Monograph Series. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.

Miller, W. R., Tonigan, J. S., & Longabaugh, R. (1995). The Drinker Inventory of Consequences (DrInC): An Instrument for Assessing Adverse Consequences of Alcohol Abuse. Project Match Monograph Series, Vol. 4. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.

Sobell, L. C., & Sobell, M. B. (1986). Can we do without alcohol abusers' self-reports? The Behavior Therapist, *7*, 141-146.

Sobell, L. C., Sobell, M. B., & Ward, E. (Eds.). (1980). Evaluating alcohol and drug abuse effectiveness. New York: Pergamon Press.

Participants in the panel discussion were Raymond Anton, Joanne Fertig, Raye Litton, Stephanie O'Malley, and Damaris Rohsenow.

3. OUTCOME MEASURES FOR COCAINE ABUSE

Chair: Deborah Leiderman
Rapporteur: George Bigelow

This group discussed four questions. The questions and a summary of the discussion are provided below. A recurring theme in the discussion was that the anti-cocaine pharmacotherapy field can benefit from, and perhaps should be guided by, clinical research experience with other effective substance abuse therapies -- with effective psychosocial/behavioral therapies, and with the effective pharmacotherapies for opioid, alcohol, and nicotine abuse and dependence. Those bodies of clinical research demonstrate procedures that have worked in detecting therapeutic efficacy and in persuading the scientific and regulatory communities that the efficacy is of a clinically meaningful magnitude. Similar procedures seem likely to be appropriate in the assessment of anti-cocaine pharmacotherapies.

a. What should be the primary outcome measures in a pharmacotherapy trial for cocaine abuse? What secondary measures should be obtained, and how should these be used in light of the primary measures?

The cocaine abuse pharmacotherapy field is disadvantaged in this discussion by the fact that there are no proven effective anti-cocaine pharmacotherapies to which we can look to look for guidance. If there were such positive examples they could provide guidance about the relative sensitivity of different outcome measures, and about their relative scientific/clinical acceptance as representing “meaningful differences” in outcome. In the absence of such specifically-relevant guidance, the group felt the anti-cocaine pharmacotherapy field should be guided by the successful experiences with psychosocial/behavioral treatments and with pharmacotherapies for opioid, alcohol, and nicotine abuse and dependence.

The group felt that the search for effective anti-cocaine pharmacotherapies has been impeded by an excessive reliance upon urinalysis toxicology data as the primary -- and often the only -- outcome measure. Urinalysis toxicology results are a relatively insensitive index of changes in the drug use behavior they are intended to reflect. Although toxicology testing can be very sensitive for detecting the occurrence of drug use in usually-abstinent individuals, it is not sensitive in detecting either reductions in drug use or periods of up to 2-3 days abstinence in individuals who continue some use. The consequence has been that, with many of the populations included in anti-cocaine pharmacotherapy trials, extremely large reductions in use (e.g., perhaps up to 90% reductions in use) might be required before even a modest reduction in urinalysis-positive results could be expected. Consequently, the group felt there should be substantially greater reliance on and acceptance of self-reported-use data in anti-cocaine pharmacotherapy trials. The group noted that self-reported-use data have been well-accepted and widely used in clinical trials of treatments for alcohol and/or nicotine dependence. Despite concerns about the limitations of urinalysis data alone, the group certainly felt that any trial that found significant treatment effects using urinalysis data alone would be very powerfully convincing.

The consensus of the group was that the best overall outcome measure was a composite index of abstinence derived from a combination of confidential patient self-report and objective

biological testing (typically urinalysis testing). The recommendation was that this composite index of abstinence be used to classify each day as abstinent or non-abstinent and that the primary outcome analyses be based on these classifications. The group felt that the confidential patient self-report component of the composite should be “default” data, only being overruled and replaced when urinalysis data contradict and disconfirm self-reported abstinence.

The group cautioned that it is essential that self-reported use data be collected in a manner in which patients are assured of the data’s confidentiality from treatment staff. Self-reported-use data collected by or not clearly confidential from treatment staff carry the risk of being seriously biased or distorted by the expectations and contingencies inherent in the treatment relationship. In contrast, numerous studies comparing self-report and urinalysis data have shown that self-reports have good validity when collected with confidentiality assured.

The composite abstinence data could be used to evaluate outcomes in a number of different ways. One might compare treatment groups on the number or percent of days abstinent, on the percentages of patients abstinent at successive time points, on the percentages of patients continuously abstinent for various durations, on the durations of longest continuous abstinence, on the time elapsed to achieve some criterion duration of abstinence, on the time elapsed to meet some criterion definition of use or relapse, and no doubt there are other indices that could be derived from these abstinence classification data. The committee offered no strong recommendation about which specific derived index might be best. Some argued in favor of the longest duration of continuous abstinence as a sensitive and clinically meaningful index. Others felt it was not ideal. The consensus seemed to be that real and meaningful therapeutic effects would likely be detectable with a variety of different specific indices. It was suggested that NIDA might benefit by arranging for comparative reanalysis of data from effective therapeutic trials (effective psychosocial/behavioral treatments, and effective pharmacotherapies for opioid, alcohol, and nicotine abuse and dependence) that would compare the relative sensitivities of different specific outcome analyses for detecting treatment efficacy.

A variety of useful and relevant secondary measures were discussed. Three general categories were discussed -- items from the Addiction Severity Index, indices of craving, and clinician ratings. All were included in the secondary measures category because it was felt that treatment effects on these indices alone -- without concurrent effects on drug use itself -- would be unconvincing, and unlikely to be adequate basis for regulatory approval or for clinical acceptance.

One problem faced by the drug abuse treatment field is the stigma of associated with drug abuse and drug abusers. One consequence is that it is desirable to include in clinical trials some indices that can reflect the societal benefits of drug abuse treatment. Items from the Addiction Severity Index such as Days Employed, Income From Employment, Days of Illegal Activity, Income From Illegal Activity were thought useful for addressing this need as well as for providing a broader picture of the lifestyle change benefits that patients achieve in treatment. These specific items were thought to be more useful and informative -- and probably more statistically sensitive -- than the composite scale scores that can be derived from the Addiction Severity Index.

There were mixed opinions about the possible value of clinician ratings of global improvement. Some felt they could be valuable and sensitive indices that integrated a broad array of data into a single composite outcome index. Others were more skeptical. While clinician ratings do have a proven track record in some fields, that is not the case in the field of

substance abuse treatment. In the absence of experience and data indicating the sensitivity and value of clinician ratings for detecting drug abuse treatment efficacy, the group felt that clinician ratings should probably not be a primary outcome variable, but could be valuable and informative to include as a secondary outcome variable.

b. What are realistic treatment goals for treatment of cocaine abuse? What weight should be given to reduction in use, as distinct from abstinence? How can use or lapse be distinguished from relapse in a clinical trial?

The group felt that all three of the major treatment goals generally discussed for cocaine abuse treatment -- cessation of use, prevention of relapse, and reduction of use -- were viable, appropriate, and clinically meaningful. All three goals were deemed to be reasonable and realistic and worthy of pursuit.

The group felt that reduction of use was perhaps a more difficult outcome for which to make a persuasive case that treatment was producing a clinically meaningful behavior change. In particular, the group felt that meaningful reductions in use should be reflected in increased days of abstinence. Reductions in use per day without reductions in days of use were thought not to be very persuasive of real benefit. In particular, it was also felt that interpretation of results from clinical trials reporting reduced use would require supplemental data documenting that the putative pharmacotherapy had not simply potentiated cocaine's effects, since such potentiation could yield reduced amounts of use.

The group agreed that there is no level of cocaine use that would be accepted as non-problematic -- i.e., there is no pattern of cocaine use analogous to "moderate drinking" of alcohol.

At this time there is no accepted or standard definition differentiating lapses from relapses. Investigators may create their own definitions for purposes of clinical trial analyses. Further research on the natural history of cocaine use and cocaine abuse treatment may provide an empirical answer to whether this is a meaningful distinction and, if so, how best to define it.

c. Should every pharmacotherapy trial include a psychosocial component? If so, how can this be standardized?

Simply at a practical level it was felt that all or nearly all trials should include a psychosocial component. Some psychosocial component is probably needed simply to engage and structure volunteers' participation. In addition, ethical considerations in most settings would require an acceptable "usual care" psychosocial component -- especially since many pharmacotherapy trial participants would likely receive only placebo pharmacotherapy.

Beyond these practical and ethical considerations there is no empirical guidance available about what the "right" level or type of psychosocial therapy is. It is unknown whether pharmacotherapy effects are most sensitively revealed in the context of minimal psychosocial care or in the context of intensive psychosocial care. Available data from nicotine dependence studies suggests the the pharmacotherapy "effect size" is similar across a very broad range of psychosocial care intensities (from minimal intervention advice to weekly individualized counseling). The relationship of psychosocial treatment intensity to pharmacotherapy effect size is a topic deserving substantially more empirical investigation. But it is a topic to which the anti-cocaine pharmacotherapy field can contribute little as yet. These studies need to be conducted with known effective pharmacotherapies.

There was consensus that pharmacotherapy trials need to standardize the concurrent psychosocial/behavioral intervention in some way. Psychosocial/behavioral interventions certainly have efficacy and can influence outcomes. They can not be permitted to vary on an “as needed” basis for individuals within a pharmacotherapy trial (though one could imagine a trial in which the extent of “as needed” psychosocial treatment was used as an outcome index of the efficacy of a pharmacotherapy). The usual model will be one in which the efficacy of the psychosocial treatment should be held constant while the pharmacotherapy is varied.

The two major dimensions in which the efficacy of the psychosocial treatment should be standardized are content and amount. There was consensus that pharmacotherapy trials should provide a manual describing the content of the psychosocial/behavioral treatment and that they should specify, control, and measure the amount of treatment provided. Some investigators thought that the potential efficacy variations among different psychosocial treatments was so great that it was important to record and review session content to ensure adherence to the prescribed psychosocial treatment. Other investigators noted that a very broad range of different manualized psychosocial “talk” therapies appear to have equivalent efficacy and felt that monitoring of adherence to a detailed prescription could be unnecessary. All agreed that some specification of the content of the psychosocial treatment and some control and specification of its delivery should be included in pharmacotherapy trials. The differences of opinion related to how precise these specifications need to be to avoid significant variation in the efficacy of the psychosocial component.

d. What are the most appropriate statistics to evaluate outcome, and how should dropouts be handled?

The group devoted little discussion to this topic, preferring instead to defer largely to the separate discussion group specializing in statistical/analytical recommendations. However, there was certainly a feeling that the difficulties experienced in developing anti-cocaine pharmacotherapies are not due to inadequate statistical/analytical procedures. The success in developing and gaining FDA approval of pharmacotherapies for other types of substance abuse and dependence (opioids, alcohol, nicotine) and the success in demonstrating the effectiveness of psychosocial/behavioral treatments demonstrate that adequate statistical/analytical procedures are available when effective therapies are available for testing.

The group felt there was no single statistical/analytical strategy to be recommended, but rather that a variety of different techniques are likely to be adequate. The group felt that guidance should be taken from the various techniques that have proven successful in analyzing other effective pharmacotherapies (for opioids, alcohol, nicotine) and for analyzing effective psychosocial/behavioral therapies. In those areas -- where there are proven and effective therapies -- it would be possible to perform comparative evaluations of the relative sensitivity of different analytic strategies. It might well be a worthwhile endeavor for NIDA to arrange for the comparative reanalysis of data from effective therapies to answer empirically whether there are consistent advantages to particular analytic approaches. The group seemed to feel that if there are real and clinically meaningful therapeutic effects they should be detectable with a variety of different analytic approaches.

The group agreed with the recommendation of the statistics/analysis discussion group that great attention should be devoted to minimizing dropouts from data acquisition. Even if patients drop out of treatment there should still be extensive effort to collect outcome data. At

the same time, the group felt that the task in this regard for assessing anti-cocaine pharmacotherapies should not be made any more stringent than is the case for assessing other types of pharmacotherapies or psychosocial therapies.

The group was skeptical of data-imputation procedures such as the “last observation carried forward” approach which could fill in imputed values for substantial missing data. Certainly the extent of missing data and data imputation needs to be reported, and substantial imputation raises serious questions about the validity of reported results. The group seemed to prefer procedures that avoid imputation and simply report the extent of observed abstinence.

4. STATISTICS IN ASSESSING OUTCOME

Chair: Peter Bridge

Rapporteur: P. Lavori

This group focussed most of its discussion on studies of cocaine pharmacotherapy. The statistics group emphasized the critical importance of complete collection of data on all outcomes in all randomized patients, regardless of changes in adherence to the study treatment protocol. The intention-to-treat analysis depends on the full follow-up of outcomes. Furthermore, methods that purport to deal with missing data (such as Generalized Estimating Equations, Random Regression, and Multiple Imputation) all rely on assuming 'Missing at Random' (MAR). This amounts essentially to assuming that patients with the same trajectory of observed outcomes have the same chance of missing a datum (such as the BE level of urine) regardless of the value of that datum (such as whether it is positive or negative for cocaine). This is highly unlikely to be true in the context of cocaine trials. Furthermore, since the assumption is not verifiable in the observed data, it is prudent to avoid such high rates of missingness that the entire inference rests upon an untestable assumption.

It is possible to 'define away' this problem if one is willing to entertain the approach that counts a missing datum in a pre-specified way (for example, treats missing urine as positive). The group emphasized that this is a clinical solution, not a statistical one, and stands or falls on its merits as a useful summary of the patient's outcomes. Such a re-definition of outcome may be questioned, on substantive grounds. There are some contexts (and patients) for whom the cessation of treatment and measurement may follow from a successful outcome (for example, in nicotine addiction). As attractive as this approach may be for finessing the measurement issue, it may not stand up to critical scrutiny. Thus, the group recommended that all reasonable attempts be made to obtain full follow-up.

The group chose to address the specific task of identifying statistical approaches to the conduct of relapse studies. This was seen as a new opportunity in the assessment of potential cocaine pharmacotherapies. The design of a relapse study in cocaine abuse appears to offer the best prospect for implementing essentially complete follow-up. Because the statistical group met without the clinical group(s), there was no effort to define 'relapse'. Return to use may comprise anything from a single use (lapse) to a binge. For the purposes of discussion relapse was taken as synonymous with a positive urine in a patient who is randomized in an 'abstinent state'. This relapse context is simpler because all that is necessary to achieve perfect measurement of the time to relapse is to observe all urine up to and including the first dirty urine, in each patient who 'relapses'. The missingness will then occur as an initial missed urine in a patient who is previously known to be abstinent. The group suggests that at the time that such a patient misses a urine, the investigator begins special procedures, whose intent is to obtain retrospective estimates of the true state of cocaine use of the patient at the time of the missing urine. These measures could include a patch or hair sample, or self report of use, if that is considered sufficiently reliable. The group noted that it was not necessary for the method to be as precise as the standard (BE) assay, just that it be unbiased. If it is a noisy (but unbiased) measure, then the statisticians will be able to estimate the necessary sample size adjustments to compensate for the extra variability, and the problem would be much more tractable. A longer

stretch of missing data would be handled in the same way. It may only be possible to ascertain that a patient has relapsed at some unknown time between the last (negative) observed urine and the subsequent successful measure of cocaine residue in hair (for example). It would not be necessary to precisely pinpoint the time in order to obtain unbiased estimates of survival probabilities, under relatively weak assumptions about the distribution of times (it is essentially a non-informative left censoring problem). Finally, even biased methods of ascertainment may produce useful information, if the degree of bias is not large and if it is estimable from calibration studies. As long as the bias does not vary across treatments, one may obtain good estimates of treatment effects with moderate bias.

Interpretation of time to relapse will vary dependent upon the anticipated approach statistically for analysis. If the contrast is to be between arms of a study using grouped data, then the pattern of prior use, including the presence of prior periods of abstinence and their duration, is principally an issue of stratification or randomization. Self report data would seem appropriate as the basis for ascertaining that by randomization all the individuals with a history of extended periods of abstinence prior to relapse were not randomized to one of the study. More recently, there has been an interest in case based analyses, categorizing individuals as successful based on prospective definitions of clinical outcome. Under a case based approach, knowledge of usage pattern may require direct observation rather than self report. This will need to be a determination of the clinicians who direct this study and those who evaluate it.

One way to think about these adjunctive methods of measurement is that they provide auxiliary data that can be used to improve the plausibility of the MAR assumption. That is, non-MAR missingness may be closer to MAR after conditioning on the values of auxiliary variables bearing on the missing target observation (e.g., the value of a cocaine assay from a specimen of hair taken a short time after a missing urine, in a hitherto abstinent patient). Furthermore, the relationship between the observed auxiliary measures can be used to reduce the lost information fraction (used as a term of art), that results from a given loss of directly observed outcomes. This may reduce the scope of the problem to the point that sensitivity analyses show that the main conclusions do not depend qualitatively on untestable MAR assumptions.

With these ideas in mind, the group believed that it might be possible to achieve very low residual levels of missing relapse data (corresponding to no more than 10% or so missing information) over moderate lengths of study time (6 months or more). If the levels of missingness can be brought down to those levels, then the residual missingness may be amenable to one of the standard missing data techniques, with a sensitivity analysis for non-ignorability.

The group also discussed the question of when to 'drop a patient from study', and strongly recommended that this not occur until the patient had demonstrated a relapse, and therefore would have provided the outcome. In particular, a patient who has fallen away from the treatment protocol but has remained well should be encouraged to return to the protocol whenever possible. This makes sense statistically and scientifically, under the intent to treat principle, which contrasts the strategies embodied in the randomization. A strategy that permits patients to re-enter as long as they remain well is likely to be a stronger one than a strategy that prohibits re-entry, since in the latter case avoidable relapses will be counted against the strategy. It will be a study design choice as to whether individuals who resume study participation after a lapse are part of the initial protocol, or are referred to a second, but linked protocol for separate analysis.

The issue of power came up, and it was acknowledged that the survival paradigm is intrinsically less powerful than a scalar comparison, unless the study lasts long enough that nearly all patients relapse. However, this comparison only favors scalar experiments that are fully observed. As soon as any substantial amount of missing data occurs (especially non-MAR) in a scalar study, the balance of power favors the fully observed relapse study. If we can make a proportional hazards assumption, then the power of the survival analysis depends only on the effect size (log hazard ratio of treated to control- LHR) and the total number of relapses, in a rather simple way. For example, to detect a halving of the relapse hazard with 80% power and a 2-sided logrank test of significance set at 5%, one needs to observe a total of about 64 relapses. If the average rate of relapse over some period of time about 1/3 then about 200 patients should be randomized (100 per group). This calculation can be elaborated to take account of periods of accrual and also (importantly) to provide estimates of the adjustments needed to accommodate both non-compliance to active treatment and non-compliance to control (idrop-inî). Such a procedure will help to predict the impact of treatment noncompliance on the results of the intent to treat analysis, and if undertaken as the study is planned, will ensure that the study is not grossly underpowered for realistic alternatives.

The group discussed whether it was permissible to cease follow-up measurement after the patient relapses. There is no statistical necessity for doing so, but given the arbitrariness of relapse definitions in this area, it might be useful to be able to characterize the depth and severity of the relapse. This is particularly true if the relapse is measured indirectly (as described above). If the patient has an inferred relapse at week 15, and then is observed to have an episode of cocaine use at week 18, the analysis will be relatively insensitive to a mistake in the earlier relapse attribution. Thus, with a combination of special effort and continued measurement, it may be possible to resolve much of the uncertainty about the status of the patient.

The extension of these ideas to the acute treatment phase (induction of abstinence) seems more difficult. However, there is no comfort in the fact that it is hard to get full measurement in that context, since the difficulties of inference are just as severe (or worse). It would seem to be prudent to avoid conducting experiments that result in uninterpretable large missingness rates, even if it is possible to do so.

Because relapse studies have not been evaluated extensively, if at all, in cocaine pharmacotherapy, it was recommended that estimations of inaturalisticî relapse rates in the context of clinical trial conduct be evaluated. Data sets from prior placebo controlled clinical trials which permit participants to be randomized with urines negative for Benzoyllecognine can be analyzed for relapse to use rates under placebo administration conditions. This pattern may vary considerably from any individualsí resumption of cocaine use following the observation of urines negative for BE when not be observed under clinical trials conditions, but it will provide some assessment of a placebo based effect against which any active treatment will be measured.

Dependent upon the frequency of return to use events in these placebo controlled study samples, and upon the duration and adequacy of observation subsequent to relapse event(s), it may be possible to characterize the parameter of relapse in the clinical population. Examples might include distribution of single compared to multiple usage episodes; the likelihood of missing data, and /or study dropouts following relapse. Because of access to and across data sets necessary for this task, it would seem most feasible if NIDA were to initiate these analyses.

5. FOLLOWUP ISSUES IN PHARMACOTHERAPY

James R. McKay

Most studies that have examined the efficacy of various pharmacotherapy agents for the treatment of substance abuse have not followed up patients beyond the end of the medication trial (usually 12 weeks). One of the justifications for not conducting posttreatment follow-ups is that treatment response within the first 90 days is highly predictive of longer-term outcomes (Carroll et al., 1994; McKay et al., 1999). Furthermore, posttreatment follow-ups are expensive and can delay the reporting of study results.

However, it may be advantageous to conduct posttreatment follow-ups of six months duration with medications that demonstrate efficacy during a trial. First, the follow-up period would allow for a determination of how long the medication effect persists after discontinuation (O'Malley et al., 1996). This information would greatly facilitate the development of recommendations for the length of time patients should stay on a medication. Second, certain medications may interact with the type of behavioral therapy that is concurrently provided, leading to the emergence of treatment effects in the posttreatment period (Carroll et al., 1994). For example, a medication might relieve distress or craving for long enough to create a "window of opportunity," during which a behavioral treatment with more lasting effects could take hold. The only way to detect such medication x behavioral treatment interaction is to follow patients beyond the end of the medication trial. In general, posttreatment follow-ups may be most appropriate for medications that are intended either to directly eliminate or reduce addiction for an extended period of time (e.g., cocaine vaccine) or to facilitate the impact of other interventions with lasting effects. Conversely, long-term follow-up may not be as useful with medications that are intended to be beneficial only for as long as they are taken (e.g., antihypertensives). Finally, it is strongly recommended that vigorous attempts be made to obtain biological and self-report outcome data from all subjects, including study dropouts, at the end of the medication trial and at any posttreatment follow-ups.

References

- Carroll, K.M., Rounsaville, B.J., Nich, C., Gordon, L.T., Wirtz, P.W., Gawin, F.H. (1994). One year follow-up of psychotherapy and pharmacotherapy for cocaine dependence: Delayed emergence of psychotherapy effects. *Archives of General Psychiatry*, 51, 989-997.
- McKay, J.R., Alterman, A.I., Cacciola, J.S., O'Brien, C.P., Koppenhaver, J.M., & Shepard, D.S. (1999). Continuing care for cocaine dependence: Comprehensive 2-year outcomes. *Journal of Consulting and Clinical Psychology*, 67.
- O'Malley, S.S., Jaffe, A.J., Chang, G., Rode, S., Schottenfeld, R.S., Meyer, R.E., & Rounsaville, B.J. (1996). Six month follow-up of naltrexone and psychotherapy for alcohol dependence. *Archives of General Psychiatry*, 53, 217-224.

Appendix: Meeting participants

Raymond F. Anton, M.D.
George E. Bigelow, Ph.D.
Dan Bloch, Ph.D.
Peter Bridge, M.D.
Harriet deWit, Ph.D.
Dorothy Hatsukami, Ph.D.
Stephen T. Higgins, Ph.D.
John R. Hughes, M.D.
Dr. J.P. Hsu
Andy Johnston, Ph.D.
Thomas R. Kosten, M.D.
Grant Ko, MD
Philip Lavori, Ph.D.
Robert S. Mansbach Ph.D.
Cynthia G. McCormick, M.D.
James McKay, Ph.D.
Stephanie O'Malley, Ph.D.
Kenneth Perkins, Ph.D.
Thomas Permutt, Ph.D.
Damaris J. Rohsenow, Ph.D.
Richard S. Schottenfeld, M.D.
Celia J. Winchell, M.D.
Jane Acri, Ph.D.
Jack D. Blaine, M.D.
Nora Chiang, Ph.D.
Acting Chief,
Richard Hawks, Ph.D.
Susan Herbert, R.N., M.S.
Jim Hill, Ph.D.
Deborah Leiderman, M.D.
Shou-Hua Li, Ph.D.
Dorota Majewska, Ph.D.
David McCann, Ph.D.
Ann Montgomery, R.N., M.S.
Lisa Onken, Ph.D.
Kenzie Preston, M.D.
Steve Sparenborg, Ph.D.
Betty Tai, Ph.D.
Frank Vocci, Ph.D.
Robert Walsh, B.S.
Stephen R. Zukin, M.D.
John Allen, Ph.D.
Joanne Fertig, Ph.D.
Richard K. Fuller, Ph.D.

Raye Litten, Ph.D.
Margaret Mattson, Ph.D.