Determining the Primary Endpoint for a Stimulant Abuse Trial: Lessons Learned from STRIDE (CTN 0037)

Madhukar H. Trivedi, M.D.[1], **Tracy L. Greer, Ph.D.**[1], **Jennifer Sharpe Potter, Ph.D., M.P.H.**[2], **Bruce D. Grannemann, M.S.**[1], **Edward V. Nunes, M.D.**[3], **Chad Rethorst, Ph.D.**[1], **Diane Warden, Ph.D., M.B.A.**[1], **Kolette M. Ring, B.A.**[1], and **Eugene Somoza, M.D., Ph.D.**[4]

[1]Department of Psychiatry, The University of Texas Southwestern Medical Center at Dallas, Dallas, Texas, 75390-9119, USA

[2]Department of Psychiatry, University of Texas Health Science Center at San Antonio, San Antonio, Texas, 78229, USA

[3]Columbia University College of Physicians and Surgeons, New York State Psychiatric Institute, New York, NY 10032, USA

[4]Cincinnati VA Medical Center, and Department of Psychiatry, University of Cincinnati/CinARC, Cincinnati, OH 45220, USA

Abstract

Background—No consensus is available for identifying the best primary outcome for substance abuse trials. While abstinence is the most desirable outcome for substance use interventions, a wide variety of other endpoints have been used to evaluate efficacy trials.

Objectives—This report provides a framework for determining an optimal primary endpoint and the relevant measurement approach for substance use disorder treatment trials. The framework was developed based on a trial for stimulant abuse using exercise as an augmentation treatment, delivered within the NIDA Clinical Trials Network. The use of a common primary endpoint across trials will facilitate comparisons of treatment efficacy.

Methods—Primary endpoint options in existing substance abuse studies were evaluated. This evaluation included surveys of the literature for endpoints and measurement approaches, followed by assessment of endpoint choices against study design issues, population characteristics, tests of sensitivity and tests of clinical meaningfulness.

Conclusion—We concluded that the best current choice for a primary endpoint is percent days abstinent, as measured by the Time Line Follow Back (TLFB) interview conducted three times a week with recall aided by a take-home Substance Use Diary. To further improve the accuracy of the self-reported drug use, an algorithm will be applied to reconcile the results from the TLFB with the results of qualitative urine drug screens.

Scientific Significance—There is a need for a standardized endpoint in this field to allow for comparison across treatment studies, and we suggest that the recommended endpoint be considered for use in this field.

Keywords

cocaine abuse; methamphetamine abuse; measurement; abstinence endpoint; exercise

Introduction

Stimulant abuse is a chronic, relapsing illness with few highly efficacious treatments. Research in the treatment of substance use disorders (SUDs) suggests a need for innovative treatments as a substantial proportion of individuals engaged in currently available treatments do not achieve abstinence (1). Research devoted to treating individuals with stimulant abuse and dependence has relied on a variety of endpoints and it is challenging to compare trial outcomes across treatment studies. Long-term continuous abstinence is the preferred clinical goal for individuals being treated for stimulant use disorders. However, the lack of efficacious treatments can make long-term continuous abstinence an unrealistic goal for research studies evaluating treatments (2–5). Furthermore, inherent weaknesses of both self-report and biological outcome measures make the evaluation of long-term continuous abstinence difficult (6, 7). Therefore, long-term continuous abstinence has most often not been seen as the best endpoint to select in the context of a clinical trial. Nonetheless, an endpoint serving as a proximate outcome must be selected that is consistent with the modest efficacy of currently available treatments.

Research in this field has typically chosen endpoints that are clinically believed to be meaningful proxy measures or intermediate endpoints toward the goal of long-term continuous abstinence. These endpoints have included a reduction in use, increases in days/ weeks of abstinence, total amount of money spent, or continuous abstinence over specified periods of time (8). Likewise, specific measurement approaches to define study endpoints also vary widely, ranging from self-reported rates of abstinence to detailed toxicology data that quantify exact amount of drug use. Occasionally functional measures have been used, as well as measures evaluating the impact of substance use on society. Currently, there is no gold standard primary endpoint or outcome measure in the field for trials of stimulant abuse treatment. Investigators who are developing trials to assess substance use outcomes must consider several important factors when deciding upon a primary endpoint and relevant outcome measure (9).

This report describes the framework used and issues involved in selecting a primary endpoint and primary outcome measure for the CTN 0037 trial, Stimulant Reduction Intervention using Dosed Exercise (STRIDE), a trial to be conducted through the Clinical Trials Network of the National Institute of Drug Abuse. Hughes et al. (10) previously reported the approach they used to help move the field toward a consensus approach to the study of smoking cessation treatments. This report aims to provide a similar contribution for the field of stimulant use. We present the model we used as well as steps undertaken to select our primary endpoint and outcome measure. Lessons learned through the completion of these steps for STRIDE are also reported.

Procedure to Determine an Optimal Endpoint for a Study

The choice of a primary endpoint is driven by the intent of the study, options in the existing literature, and consistency of the choice with features of the study design (9, 11, 12). The first step in selecting a primary study endpoint is to conduct a review of the literature to determine the available endpoints used within the specific population being studied, as well as the endpoints specifically relevant to the intervention. Once the candidate endpoints are determined, the next step is to select the endpoint that best fits the intent of the study. Finally, it must be determined whether the selected endpoint is clinically meaningful. While these procedures are described as if they occur in a linear fashion, this is an iterative process with ongoing reevaluation of critical factors and decisions by experts in the interventions and disorders of interest. Figure 1 presents our process for deciding upon a primary endpoint for the STRIDE trial.

Identify the Intent of the Study

STRIDE is a randomized, controlled trial that aims to assess the addition of vigorous-intensity, high-dose exercise or health education to treatment as usual (TAU; i.e., usual care) in improving stimulant use disorder treatment outcomes. Individuals diagnosed with stimulant abuse or dependence (cocaine, methamphetamine, amphetamine, or other stimulant, except caffeine or nicotine) will receive 3 months of acute phase supervised intervention followed by an additional 6 months of supervised intervention once per week. Both groups will receive stimulant abuse TAU, which will begin in a residential setting, followed by community treatment.

A Protocol Development Team was assembled for the trial that was comprised of experts in stimulant abuse treatment and research, exercise and behavioral interventions, and multisite clinical trial research (Appendix 1). In addition to the overall development of the study design and protocol, this team was tasked with identifying the best study endpoint to evaluate whether or not exercise or health education enhances usual care for stimulant use disorders.

There are a number of important outcomes appropriate to consider when investigating SUD treatments (e.g., psychosocial, health-related outcomes) (13–18). As this is an early efficacy trial, the study primary outcome focuses on the more immediate substance use outcomes rather than other, often longer term problem related outcomes with possibly lower effect sizes. Problem related outcomes will also be collected, however.

Review of the Literature for Candidate Endpoints in Use

We reviewed the existing literature to determine if a standard endpoint was available and suitable for this population (stimulant abusers), setting (residential treatment followed by outpatient care) and the interventions being studied (exercise and health education). Given that STRIDE is a study of a novel treatment approach for stimulant abusers with an innovative study design that follows a participant through a treatment process that includes both residential and outpatient care, no standard endpoint was available.

Because there was no clear choice of endpoint based on this search, we then searched for studies of other behavioral interventions that targeted SUDs in general and those that specifically targeted stimulant abuse, as well as studies examining exercise interventions in other relevant clinical populations (e.g., exercise augmentation for smoking cessation).

Again, endpoints in outcomes research in SUDs are variable, and range from abstinence, to reduction in drug use over time, to relapse prevention (8). A meta-analysis of trials

examining psychosocial interventions for SUDs found that abstinence, either self-report, based on urine drug screen, or a combination, was indeed a common endpoint in SUD trials, although several different definitions have been used for abstinence, including mean weeks (or days) of abstinence, mean percent of weeks (or days) abstinent, percent of sample abstinent for a particular duration of time, percent of sample achieving post-treatment abstinence, and post-treatment scores on the Addiction Severity Index drug scale. Time to relapse has also been used, with varying measures of relapse, such as time to first use after abstinence, or time to second or third use (1). Finally, time to treatment discontinuation and treatment retention have also been used; these approaches generally assume that attrition indicates resumed substance use or treatment failure.

Focusing specifically on treatment studies for stimulant users, the variability in endpoints was again apparent. Endpoints used included continuous abstinence (11, 12, 19, 20), mean percent or duration of abstinence (21, 22), treatment retention (2, 12, 23, 24), treatment attendance (12), and frequency and quantity of use (2, 3, 12, 23, 25–27). Toxicology outcomes included mean number of negative (or positive) urine screens throughout treatment and mean percent of negative (or positive) urine screens throughout treatment (7).

Finally, we searched for studies of exercise as a treatment for related clinical conditions, which yielded studies primarily focused on smoking cessation. However, no gold standard for a study endpoint emerged from a review of these trials either. These studies used a variety of study endpoints including continuous abstinence, abstinence occurring over different specific periods of time (e.g., 12 weeks, last month etc.) or at a specific point in time (e.g., 7-day point prevalence abstinence), craving and other outcomes (28–30).

Selection of a Candidate Endpoint

Based on our review of the literature, it was clear there were several endpoints we could consider for STRIDE. Table 1 illustrates the endpoints that were considered by the STRIDE Protocol Development Team, and presents the advantages and disadvantages of each.

Because it is most representative of long-term continuous abstinence, we first considered consecutive days or weeks of abstinence for our primary endpoint. However, given that this study examines a novel augmentation treatment for stimulant use disorders in a field where augmentation strategies offer only modest incremental improvements (e.g., Carroll et al. (31) reported that the effect size for active psychotherapies compared with a control condition was d=0.16), we decided it was more meaningful to choose an endpoint that would also identify reduction in use. For example, if a participant entered the study after typically using stimulants 3–4 days per week, but reduced usage to once per week or once every other week, consecutive days/weeks of abstinence would not adequately capture that reduction. We therefore decided that percent days of abstinence was the most reasonable primary endpoint for STRIDE.

Selection of a Primary Outcome Measure

After identifying a primary endpoint, we next had to establish our definition of an abstinent day. We again began with a search of relevant literature. The most common outcome measure utilized in treatment studies for stimulant abuse and/or dependent participants was urine drug screens, typically collected 1–3 times per week. However, studies varied widely with respect to the use of self-report or objective measures of drug use, with some trials using only one or the other, and other trials using a combination of the two. Because no consistent outcome measure has been used to assess stimulant use, we assessed the strengths and weaknesses of potential outcome measures. The main options considered by the Protocol Development Team (PDT) are detailed in Table 2.

After considering the potential advantages and disadvantages, we decided upon an approach that integrates the strengths of multiple measures and increases the chances of successful data collection over an extended time period. We chose to assess percent days abstinent during the observation period as measured by self-report using a Time Line Follow Back (TLFB) (32). The TLFB interview is conducted in a nonjudgmental manner by trained research staff, with no adverse consequences for disclosure of use. Participants are informed that their disclosed information is not shared with others, such as clinic staff or treatment providers. This is a critical feature of this process, since individuals are likely to be unwilling to disclose substance use in situations where their report of use would have severe repercussions (e.g., a parolee reporting use to their probation officer). The TLFB has been shown to correlate well with objective measures of use such as urine drug screen (33–35). As noted in Table 2, disadvantages of the TLFB include potential inaccuracy due to memory errors and bias and deliberate and denial-based distortions of reported SUD. To further bolster the accuracy of our assessment, we added two additional tools to the data collection process. First, the participants will use a Substance Use Diary to aid with recall during the TLFB assessment. The Substance Use Diary is a calendar that participants are asked to take home to record use in real time, as well as document other events (e.g., birthdays, travel) that can be helpful to trigger recall during the TLFB interview. The diary is intended to be used in between study visits to help participants maintain an accurate accounting of their usage of stimulants and other drugs. It can be particularly helpful in situations where missed visits occur and greater amounts of time must be recalled. A diary is not expected to provide the same comprehensiveness of information as the structured TLFB interview, but the use of a memory aid such as a diary has been shown to improve the accuracy of the TLFB (36). Finally, to further improve the validity of the TLFB, an algorithm that includes both TLFB self-reported use as well as results from the urine drug screens collected 3 times per week will be employed at study end to reconcile discrepancies between negative TLFB reports and positive urine drug screens. The availability of the primary outcome measure is therefore not affected by an absent urine drug screen.

Design and Population Factors that Influence the Measurement of the Primary Endpoint

An additional consideration in choosing an outcome measure was the applicability of the selected outcome measure for the study design and population. In a traditional experimental design, the research question determines the study endpoint and measure, and the study is then designed around that endpoint and measure. However, applied researchers know that clinical research is highly influenced by the context in which the research will be conducted. Elements of the trial design as well as characteristics of the study population must be considered when selecting an endpoint. These design considerations include the types of outcome measures that are possible (e.g., blind vs. unblind, self-report vs. objective), the treatment setting, and the treatment duration.

Choice of the TLFB self-report measure was consistent with several elements of our study design. Our study intervention will be conducted over a 9-month period during which time participants not only transition from residential to outpatient treatment but are expected to come to the study site 3 times a week for 12 weeks and once a week for 24 more weeks. While an objective measure of stimulant use is optimal, sole reliance on UDSs is not realistic. As stimulant abusing patients are known to be inconsistent in their attendance at treatment and study visits (1) a UDS only endpoint would result in considerable missing data. Furthermore, urine drug screen results are affected by both amount and time of use. Thus, given currently available screening technology, reliance on the UDS as an exclusive outcome can be complicated by this issue as well. The TLFB, on the other hand, allows for

data to be collected on all days, since the interviewer prompts the participant to recall use for all days since the last study visit, and accuracy of the information is not impacted by amount and timing of use, as with urine drug screens.

Assessing the Clinical Meaningfulness of the Endpoint

Once the choice of outcome measure was considered with respect to design issues, the next step was to determine what magnitude of change was necessary for a clinically meaningful endpoint. The ability of potential endpoints to be readily translated and adapted for use in clinical practice must be evaluated. Continuous abstinence might be considered the ideal endpoint for individuals with stimulant abuse or dependence, but use of this endpoint has not been realistic given the time limits and expected frequency of attended visits in clinical trials. Thus, while continuous abstinence might be relevant, it is not necessarily the *best* endpoint for a particular study. Other endpoints may be clinically meaningful and have greater sensitivity and represent better endpoint choices.

Identifying the magnitude of change in a study endpoint that would be clinically meaningful can in part be determined by the literature. Often specific "cut-off" points are established in the literature to determine clinical significance. For example, Higgins et al. (11) considered a treatment to be successful if participants achieved at least 9 weeks of continuous abstinence or 92% or greater days of abstinence. Our choice of endpoint (percent days abstinent) will allow for comparison against the suggested criteria set forth by Higgins et al. Furthermore, Miller and Manuel (37) also identify other endpoints of clinical significance, including quantity of use, length of initial abstinence, use of other illicit substances, and dropout from SUD treatment. Our selection of outcome measure will allow for analysis of these secondary endpoints that may be clinically meaningful.

Another guide to the magnitude of differences that can be considered clinically meaningful are those reported in peer-reviewed literature for that field since peer reviewers must consider whether the results of studies significantly contribute to the field. However, these studies do not necessarily reflect differences that would be meaningful to clinical practitioners. Miller and Manuel (37) note that it is important to understand how clinicians view differences between treatments and whether or not a particular effect would be of interest to them or likely to change their practice. Thus, as always done in the CTN, our PDT included clinical treatment providers who provided valuable input in this regard.

Comment

In the SUD literature, there is no gold standard for a primary endpoint for treatment trials. Although continuous abstinence is the ideal outcome for treatment programs, the state of research in SUD in general and stimulant abuse in particular suggests an endpoint that allows for the modest effects of treatment interventions to be detected. This may be even more important with augmentation interventions that are adding only incremental gains.

We conducted a comprehensive evaluation of the options for primary endpoints in studies of treatments for stimulant abuse, including surveys of the literature for endpoints and measurement approaches, evaluation of the choices against study design issues, population characteristics, and tests of sensitivity and clinical meaningfulness. We have concluded that given the state of the field, the best current choice for a primary endpoint is percent days abstinent, given its ability to detect a clinically meaningful improvement in abstinence, an important clinical goal. We chose the TLFB interview with its extensive available retrospective timeframe, as the most comprehensive measure that could be collected over an extended time period and enhanced its reliability with a take-home Substance Use Diary to aid recall. Furthermore, an algorithm will be used at study end to reconcile the results

obtained from the TLFB and the objective urine drug screens. Thus, our selection of primary endpoint and outcome measure: 1) captures use information across our entire acute phase; 2) blends advantages of multiple approaches in that it is informed by collection of real-time data from the Substance Use Diary, as well as objective data from urine drug screens; 3) provides us with a clinically meaningful evaluation of the efficacy of exercise as an augmentation to treatment as usual.

Other fields (i.e., smoking cessation) have used a similar approach to help guide the field toward a consensus model of studying treatment interventions (10). Researchers may want to consider using the endpoint we have identified to improve the field's ability to compare the efficacy of interventions in trials of stimulant abuse. It will also be important for future studies to consider technological advancements in measurement capabilities, such as the use of electronic diaries and interactive voice response systems, as these will likely have a strong influence on the development of available outcomes in the future. As the field progresses in identifying better treatments for substance use, researchers may also use the framework we have described to determine an optimal primary endpoint for future studies.

Acknowledgments

References

1. Dutra L, Stathopoulou G, Basden SL, Leyro TM, Powers MB, Otto MW. A meta-analytic review of psychosocial interventions for substance use disorders. Am J Psychiatry. 2008; 165 (2):179–187. [PubMed: 18198270]

2. Carroll KM, Ball SA, Nich C, Martino S, Frankforter TL, Farentinos C, Kunkel LE, Mikulich-Gilbertson SK, Morgenstern J, Obert JL, Polcin D, Snead N, Woody GE. Motivational interviewing to improve treatment engagement and outcome in individuals seeking treatment for substance abuse: a multisite effectiveness study. Drug Alcohol Depend. 2006; 81 (3):301–312. [PubMed: 16169159]

3. Rawson RA, McCann MJ, Flammino F, Shoptaw S, Miotto K, Reiber C, Ling W. A comparison of contingency management and cognitive-behavioral approaches for stimulant-dependent individuals. Addiction. 2006; 101 (2):267–274. [PubMed: 16445555]

4. Wetzel H, Szegedi A, Scheurich A, Lorch B, Singer P, Schlafke D, Sittinger H, Wobrock T, Muller MJ, Anghelescu I, Hautzinger M. Combination treatment with nefazodone and cognitive-behavioral therapy for relapse prevention in alcohol-dependent men: a randomized controlled study. J Clin Psychiatry. 2004; 65 (10):1406–1413. [PubMed: 15491246]

5. Ball SA, Martino S, Nich C, Frankforter TL, Van Horn D, Crits-Christoph P, Woody GE, Obert JL, Farentinos C, Carroll KM. Site matters: multisite randomized trial of motivational enhancement therapy in community drug abuse clinics. J Consult Clin Psychol. 2007; 75 (4):556–567. [PubMed: 17663610]

6. Fals-Stewart W, O'Farrell TJ, Freitas TT, McFarlin SK, Rutigliano P. The timeline followback reports of psychoactive substance use by drug-abusing patients: psychometric properties. J Consult Clin Psychol. 2000; 68 (1):134–144. [PubMed: 10710848]

7. Preston KL, Silverman K, Schuster CR, Cone EJ. Assessment of cocaine use with quantitative urinalysis and estimation of new uses. Addiction. 1997; 92 (6):717–727. [PubMed: 9246799]

8. Gerstein, DR. Outcome research – drug abuse. In: Galanter, M.; Kleber, HD., editors. American Psychiatric Press Textbook of Substance Abuse Treatment. Washington, DC: American Psychiatric Press; 1999.

9. Lavori PW, Bloch DA, Bridge PT, Leiderman DB, LoCastro JS, Somoza E. Plans, designs, and analyses for clinical trials of anti-cocaine medications: where we are today. NIDA/VA/SU Working Group on Design and Analysis. J Clin Psychopharmacol. 1999; 19 (3):246–256. [PubMed: 10350031]

10. Hughes JR, Keely JP, Niaura RS, Ossip-Klein DJ, Richmond RL, Swan GE. Measures of abstinence in clinical trials: issues and recommendations. Nicotine Tob Res. 2003; 5 (1):13–25. [PubMed: 12745503]

11. Higgins ST, Budney AJ, Bickel WK, Hughes JR, Foerg F, Badger G. Achieving cocaine abstinence with a behavioral approach. Am J Psychiatry. 1993; 150 (5):763–769. [PubMed: 8480823]

12. Petry NM, Peirce JM, Stitzer ML, Blaine J, Roll JM, Cohen A, Obert J, Killeen T, Saladin ME, Cowell M, Kirby KC, Sterling R, Royer-Malvestuto C, Hamilton J, Booth RE, Macdonald M, Liebert M, Rader L, Burns R, DiMaria J, Copersino M, Stabile PQ, Kolodner K, Li R. Effect of prize-based incentives on outcomes in stimulant abusers in outpatient psychosocial treatment programs: a national drug abuse treatment clinical trials network study. Arch Gen Psychiatry. 2005; 62 (10):1148–1156. [PubMed: 16203960]

13. Allen JP. Measuring Outcome in Interventions for Alcohol Dependence and Problem Drinking: Executive Summary of a Conference Sponsored by the National Institute on Alcohol Abuse and Alcoholism. Alcohol Clin Exp Res. 2003; 27 (10):1657–1660. [PubMed: 14574237]

14. Anton RF, Randall CL. Measurement and Choice of Drinking Outcome Variables in the COMBINE Study. J Stud Alcohol Drugs. 2005; (Suppl 15):104–109.

15. Babor TF, Steinberg K, Anton R, Del Boca F. Talk is Cheap: Measuring Drinking Outcomes in Clinical Trials. J Stud Alcohol. 2000; 61 (1):55–63. [PubMed: 10627097]

16. Del Boca F, Noll JA. Truth of Consequences: the validity of self-report data in health services research on addictions. Addiction. 2000; 95(Suppl 3):S347–S360. [PubMed: 11132362]

17. Finney JW, Moyer A, Swearingen CE. Outcome Variables and Their Assessment in Alcohol Treatment Studies: 1968–1998. Alcohol Clin Exp Res. 2003; 27 (10):1671–1679. [PubMed: 14574240]

18. Miller PG, Miller WR. What should we be aiming for in the treatment of addiction? Addiction. 2009; 104:685–686. [PubMed: 19413777]

19. Silverman K, Robles E, Mudric T, Bigelow GE, Stitzer ML. A randomized trial of long-term reinforcement of cocaine abstinence in methadone-maintained patients who inject drugs. J Consult Clin Psychol. 2004; 72 (5):839–854. [PubMed: 15482042]

20. Silverman K, Higgins ST, Brooner RK, Montoya ID, Cone EJ, Schuster CR, Preston KL. Sustained cocaine abstinence in methadone maintenance patients through voucher-based reinforcement therapy. Arch Gen Psychiatry. 1996; 53 (5):409–415. [PubMed: 8624184]

21. Kirby KC, Marlowe DB, Festinger DS, Lamb RJ, Platt JJ. Schedule of voucher delivery influences initiation of cocaine abstinence. J Consult Clin Psychol. 1998; 66 (5):761–767. [PubMed: 9803694]

22. Petry NM, Martin B, Simcic F Jr. Prize reinforcement contingency management for cocaine dependence: integration with group therapy in a methadone clinic. J Consult Clin Psychol. 2005; 73 (2):354–359. [PubMed: 15796645]

23. McKee SA, Carroll KM, Sinha R, Robinson JE, Nich C, Cavallo D, O'Malley S. Enhancing brief cognitive-behavioral therapy with motivational enhancement techniques in cocaine users. Drug Alcohol Depend. 2007; 91 (1):97–101. [PubMed: 17573205]

24. Carroll KM, Rounsaville BJ, Gordon LT, Nich C, Jatlow P, Bisighini RM, Gawin FH. Psychotherapy and pharmacotherapy for ambulatory cocaine abusers. Arch Gen Psychiatry. 1994; 51 (3):177–187. [PubMed: 8122955]

25. Baker A, Lee NK, Claire M, Lewin TJ, Grant T, Pohlman S, Saunders JB, Kay-Lambkin F, Constable P, Jenner L, Carr VJ. Brief cognitive behavioural interventions for regular amphetamine users: a step in the right direction. Addiction. 2005; 100 (3):367–378. [PubMed: 15733250]

26. Somoza E, Somoza P, Lewis D, Li SH, Winhusen T, Chiang N, Vocci F, Horn P, Elkashef A. The SRPHK1 outcome measure for cocaine-dependence trials combines self-report, urine

benzoylecgonine levels, and the concordance between the two to determine a cocaine-use status for each study day. Drug Alcohol Depend. 2008; 93 (1–2):132–140. [PubMed: 18029115]

27. Carroll KM, Fenton LR, Ball SA, Nich C, Frankforter TL, Shi J, Rounsaville BJ. Efficacy of disulfiram and cognitive behavior therapy in cocaine-dependent outpatients: a randomized placebo-controlled trial. Arch Gen Psychiatry. 2004; 61 (3):264–272. [PubMed: 14993114]

28. Marcus BH, Albrecht AE, King TK, Parisi AF, Pinto BM, Roberts M, Niaura RS, Abrams DB. The efficacy of exercise as an aid for smoking cessation in women: a randomized controlled trial. Arch Intern Med. 1999; 159 (11):1229–1234. [PubMed: 10371231]

29. Martin JE, Calfas KJ, Patten CA, Polarek M, Hofstetter CR, Noto J, Beach D. Prospective evaluation of three smoking interventions in 205 recovering alcoholics: one-year results of Project SCRAP-Tobacco. J Consult Clin Psychol. 1997; 65 (1):190–194. [PubMed: 9103749]

30. Marcus BH, Lewis BA, Hogan J, King TK, Albrecht AE, Bock B, Parisi AF, Niaura R, Abrams DB. The efficacy of moderate-intensity exercise as an aid for smoking cessation in women: a randomized controlled trial. Nicotine Tob Res. 2005; 7 (6):871–880. [PubMed: 16298722]

31. Carroll KM, Nich C, Ball SA, McCance E, Rounsavile BJ. Treatment of cocaine and alcohol dependence with psychotherapy and disulfiram. Addiction. 1998; 93 (5):713–727. [PubMed: 9692270]

32. Sobell, LC.; Sobell, MB. Timeline follow-back: A technique for assessing self-reported alcohol consumption. In: Litten, RZ.; Allen, JP., editors. Measuring Alcohol Consumption: Psychosocial and Biological Methods. Totowa, NJ: Humana Press; 1992. p. 41-72.

33. Nunes EV, McGrath PJ, Quitkin FM, Ocepek-Welikson K, Stewart JW, Koenig T, Wager S, Klein DF. Imipramine treatment of cocaine abuse: possible boundaries of efficacy. Drug Alcohol Depend. 1995; 39 (3):185–195. [PubMed: 8556967]

34. Nunes EV, Quitkin FM, Donovan SJ, Deliyannides D, Ocepek-Welikson K, Koenig T, Brady R, McGrath PJ, Woody G. Imipramine treatment of opiate-dependent patients with depressive disorders. A placebo-controlled trial. Arch Gen Psychiatry. 1998; 55 (2):153–160. [PubMed: 9477929]

35. Magura S, Kang SY. Validity of self-reported drug use in high risk populations: a meta-analytical review. Subst Use Misuse. 1996; 31 (9):1131–1153. [PubMed: 8853234]

36. Hersh D, Mulgrew CL, Van Kirk J, Kranzler HR. The validity of self-reported cocaine use in two groups of cocaine abusers. J Consult Clin Psychol. 1999; 67 (1):37–42. [PubMed: 10028207]

37. Miller WR, Manuel JK. How large must a treatment effect be before it matters to practitioners? An estimation method and demonstration. Drug Alcohol Rev. 2008; 27 (5):524–528. [PubMed: 18608445]

38. McLellan AT, Luborsky L, Woody GE, O'Brien CP. An improved diagnostic evaluation instrument for substance abuse patients. The Addiction Severity Index. J Nerv Ment Dis. 1980; 168 (1):26–33. [PubMed: 7351540]

39. Cacciola JS, Alterman AI, McLellan AT, Lin YT, Lynch KG. Initial evidence for the reliability and validity of a "Lite" version of the Addiction Severity Index. Drug Alcohol Depend. 2007; 87 (2–3):297–302. [PubMed: 17045423]

40. Winhusen TM, Somoza EC, Singal B, Kim S, Horn PS, Rotrosen J. Measuring outcome in cocaine clinical trials: a comparison of sweat patches with urine toxicology and participant self-report. Addiction. 2003; 98 (3):317–324. [PubMed: 12603231]

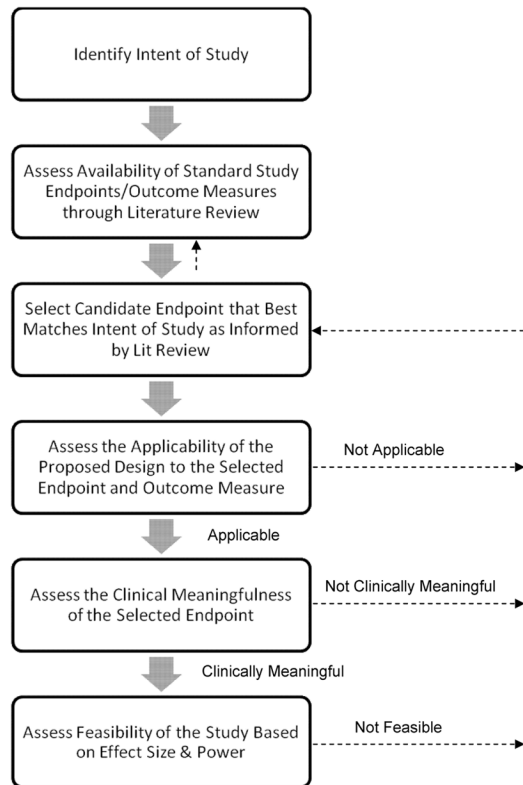41. Hawks RL, Chiang CN. Examples of specific drug assays. NIDA Res Monogr. 1986; 73:84–112. [PubMed: 3127725]

Figure 1. Flow diagram of the steps needed to decide on a primary endpoint for a clinical trial
For the STRIDE study, the intent of the study is to assess whether or not augmentation of treatment as usual with exercise or health education will improve stimulant use outcomes. Investigators, with the support and guidance of a Protocol Development Team (PDT), reviewed salient literature on treatments for substance use in general, as well as studies that used our specific interventions or that were conducted within our specific population of interest. The team decided upon a primary endpoint of percent days abstinent, as measured by the Time Line Follow Back informed by a daily Substance Use Calendar and urine drug screens collected 3 times per week. The feasibility and applicability of using the TLFB to assess percent days abstinent in a trial of long duration (i.e., 9 months) and in a setting that requires transitioning from residential to outpatient treatment was considered. Consensus was obtained by the PDT regarding the clinical meaningfulness of an endpoint of percent days abstinent. Note that if at any step a problem is encountered (e.g., power analysis reveals that the number of subjects needed to conduct the trial is not feasible, or the effect associated with the selected endpoint is deemed not clinically meaningful), one can return to the previous step(s) to reevaluate the information and select the new endpoint.

Table 1

Endpoints Considered by the STRIDE Protocol Development Team

Endpoint	Description	Advantages	Disadvantages
Number of Days Abstinent	A measure of frequency of use that can be collected by subjective report or objective measurement.	• represents a broad picture of use over any period of time and is sensitive to change in frequency of use	• ongoing abstinence is the goal of most treatment programs—this measure does not necessarily show periods of abstinence • can also miss severity of use—episodic use of greater severity may have more days of abstinence than more frequent use of less severity • if using a self-report, has potential biases and distortions of a self-report • there is no consensus regarding what decrease in use is clinically important, although effect sizes can be selected that are consistent with effect sizes in studies of other interventions
Number of Consecutive Weeks Abstinent	A measure of number of continuous weeks in which there is no use. Can be collected by subjective report or objective measurement.	• consistent with the goal of ongoing abstinence in substance abuse treatment • more stringent outcome than number of days abstinent since continuous abstinence is more difficult to achieve than intermittent days of non-use	• if using self-report, has potential biases and distortions of a self-report • may be less sensitive to change • does not report frequency or severity of use in time of non-abstinence • there is no consensus about what decrease in use is clinically important, although effect sizes can be selected that are consistent with effect sizes in studies of other interventions
Rate of Relapse	This measures percent of individuals relapsing at any particular time point. Can be measured by self-report or objective measurement.	• identifies the efficacy of a treatment at any time point • repeated measures can be used to reflect efficacy over time	• this does not measure frequency or amount of use, or timing of abstinence • provides only a snapshot of a particular time period-point prevalence
Time to Relapse	Number of days (or weeks) from discontinuation of use to first use. Can be measured by self-report or objective measurement.	• this is a measure of continuous abstinence, consistent with an abstinence goal • this is a measure typically used by people in recovery	• given the relapsing nature of substance abuse, this provides only a snapshot of the initial time period before a first relapse
Reduction in Amount of Drug Use	Change in volume of drug use. Could be based on self-report or objective measure such as quantitative urine drug screen and sweat patch.	• useful measure of severity • is relevant to a goal of use reduction rather than abstinence	• does not focus on abstinence or identify change in frequency of use • if using self-report, has potential biases and distortions of a self-report • quantitative measures can be difficult to interpret since they are affected by both time and amount of drug usage

Endpoint	Description	Advantages	Disadvantages
Reduction in Amount of Money Spent	Change in dollars spent over an identified time period.	• useful measure of severity • is relevant to a goal of use reduction rather than abstinence	• does not focus on abstinence or identify change in frequency of use • if using self-report, has potential for biases and distortions of a self-report

Table 2

Outcome Measure Options Considered by the STRIDE Protocol Development Team

Measure	Description	Advantages	Disadvantages	Psychometric Properties
Time Line Follow Back (TLFB) (26)	• Semi-structured interview based on calendar to assist in retrospective estimates • Records daily use	• self-report methods are economical, efficient, relatively noninvasive (30) • allow for evaluation of substance use over longer time periods (6) than with biological approaches • allows for reporting over an extended period of time, since patients frequently miss appointments when biological samples are taken • allows for in-depth look at substance use over time, rather than being limited to a dichotomous variable of use vs. no use • commonly used in the field and the CTN and allows for comparison between studies • can measure abstinence for days, weeks, or any other timeframe	• relies on patient report which can include memory errors and bias, with the potential for these errors to increase over longer recall periods • possibility for deliberate and denial-based distortions by the participant (6) • may be slightly less valid with longer periods of recall (6)	• High test-retest reliability (intraclass correlations range from 0.70 to 0.94) (6) • High agreement with urine drug screens (Yule's γ of 87) (6)
Substance Use Diary	• Self-report measure • Records daily drug use • Can be completed by participant at home	• could improve the accuracy of data over a TLFB interview approach since the participant completes the information in real time • advantages similar to TLFB: economical, efficient, relatively non-invasive, and providing a broader look at the substance abuse picture, compared to using biological methods	• drawbacks similar to other self-report methods: participant memory error, bias, or distortion • participant may not completely or accurately fill out the calendar without prompting from an interviewer • participant may lose the form or fail to bring it in	• Not available[#]
Addiction Severity Index-Lite (ASI-Lite) (32)	• Standardized, semi-structured interview • Assesses multiple domains that commonly affect substance abusers (32) • Can assess lifetime or current use patterns (30 days)	• used in many CTN studies, making it easier to compare results across studies • gives a view of outcome measures in the multiple domains of functioning that can be affected by substance use, in addition to amount of substance use	• length of time it takes to complete the assessment at baseline • subjective measure of drug use which relies on the participant's report • gathers estimates of substance use over time, but does not allow for specifics of quantity of use	• High intraclass correlations range from 0.79 to 0.85 for drug use (33)

Measure	Description	Advantages	Disadvantages	Psychometric Properties
			• clinical interpretation of numerical index (subscale) scores not always clear	• Not applicable *
Urine Drug Screen (UDS)	• Biological assay • Subject s are required to provide urine samples • Measurement frequency is commonly 2-3X per week • Most commonly reported as a positive or negative drug screen • Note: quantitative analyses can be conducted on urine samples	• objective measure for assessing drug use and does not rely on participant's self-report • widely used and accepted method of assessment in treatment studies (9) • results can be compared across studies (7) • well accepted approach for cocaine assessment and best-studied objective measure for diagnosing cocaine use (34)	• short detection time periods, high cost, and personal invasiveness (35) • only able to detect drug use within a limited amount of time after the use and can only show if drug use occurred, not how much drug was used or what the pattern of use was like • variance in detection sensitivity for different drugs: some drugs are detectable for longer periods of time than others, so it can miss some episodes of use • sometimes using a large amount of a substance (e.g., marijuana, cocaine) may cause more than one positive screen, even though there was only one use, referred to as "carryover" (7) • can be vulnerable to participant distortion because samples can be switched or participants can alter the sample • possibility for missing data if the participant does not show up to provide a sample	• Not applicable *
Sweat Patches	• Biological assay • Based on sweat collected from participant's skin • Good correlation between a weekly patch and patches checked 3X per week (0.69) (34) • Good correlation between patches and UDS (0.73) (34)	• can assess substance use over longer periods of time compared to a UDS, and amount of use can be determined • can reduce the variability that occurs because of the time difference between the substance use and a UDS • may also allow for the same, or in some cases greater, amount of data provided compared to a UDS, with a less invasive and potentially less expensive measure (34)	• may be more expensive than UDS • may be irritating for participants to wear • may fall off or be removed prematurely by participant, thus losing the data • there have not been as many studies completed examining sweat patch use, compared to UDS and other measures • extremely reliable for substances such as opiates, but are less reliable for stimulants	

#Psychometric properties for the Substance Use Diary have not been published.

*Psychometric properties are not relevant for biological measures.