Published in final edited form as:

Drug Alcohol Depend. 2017 December 01; 181: 219–222. doi:10.1016/j.drugalcdep.2017.09.031.

Statistical considerations in the choice of endpoint for drug use disorder trials

Garrett M. Fitzmaurice^{a,c}, Stuart R. Lipsitz^{b,c}, and Roger D. Weiss^{a,c}

^aMcLean Hospital, 115 Mill Street, Belmont, MA 02478, USA

^bBrigham and Women's Hospital, 1620 Tremont Street, Boston, MA 02120, USA

^cHarvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA

Abstract

Background—To date, the US Food and Drug Administration (FDA) requires drug use disorder trials developing new medications to use abstinence, a clinically meaningful endpoint, as the primary outcome. Although abstinence is the gold standard, only a relatively small percentage of participants in drug use disorder trials ever achieve this endpoint. This has prompted clinical trialists to consider quantitative measures of frequency of use, recognizing that some reductions in drug use that fall short of complete abstinence may potentially represent clinically important improvements. While much of the discussion concerning alternative outcomes to abstinence has focused on their clinical relevance, there are important statistical considerations that should also be taken into account.

Purpose—In this paper, we demonstrate and highlight the degree to which use of a quantitative measure of frequency of use, relative to a binary measure of abstinence, yields a very discernible increase in statistical power for assessing efficacy or effectiveness of treatments for drug use disorders.

Conclusion—While it is well established that dichotomizing a quantitative measure invariably results in loss of statistical power, what is less well recognized is the *degree* of loss in the context of drug use disorder trials. In some cases, the required sample size must be almost doubled to achieve the same level of power.

Keywords

Abstinence; Drug use; Sample size; Statistical power

Correspondence: Garrett M. Fitzmaurice, McLean Hospital, Laboratory for Psychiatric Biostatistics, 115 Mill Street, Belmont, MA 02478, fitzmaur@hsph.harvard.edu.

Contributors

All authors contributed to the conceptualization and design of the analyses. Garrett Fitzmaurice and Stuart Lipsitz conducted the analyses. Garrett Fitzmaurice wrote the first draft of the manuscript. Stuart Lipsitz and Roger Weiss provided substantive and conceptual feedback on all drafts. All authors contributed to and have approved the final manuscript.

Conflict of Interest

Roger Weiss has consulted to Indivior, Alkermes and GW Pharmaceuticals. All other authors report no conflicts of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Although the randomized clinical trial is the gold standard in study design for evaluating treatment efficacy and effectiveness, a major issue in designing trials of treatments for drug use disorders concerns the challenge of defining an appropriate endpoint. The US Food and Drug Administration (FDA) requires clinical trials developing new medications to use a period of post-treatment abstinence as the clinically meaningful endpoint (Food and Drug Administration, 2013); however, only a relatively small percentage of participants in clinical trials ever achieve enduring abstinence in the 6–12 months following completion of treatment. This has prompted interest in the development of outcome measures, other than sustained abstinence, for use as endpoints in drug use disorder trials (Carroll et al., 2014; Donovan et al., 2012; Kiluk et al., 2016; Tai et al., 1997; Tiffany et al., 2012; Van den Brink et al., 2006; Vocci and de Wit, 1999).

The use of a dichotomous (or binary) measure of abstinence, relative to a quantitative measure of *frequency of use*, entails a very discernible loss of information that in turn can result in a significant loss of statistical power (Altman and Royston, 2006; Cohen, 1983; Cox, 1957; Fedorov et al., 2009; MacCallum et al., 2002; Royston et al., 2006; Streiner, 2002). This loss of information is exacerbated by the relatively low rate of abstinence observed in many drug use disorder trials.

In this paper, we present preliminary and somewhat conservative calculations that indicate that the same statistical power can be maintained with fewer than half the number of participants when the primary outcome is a quantitative measure of frequency of use rather than abstinence; in many clinical trial settings, the reductions in sample size are even greater. Thus, from a purely statistical standpoint, there is a very strong rationale for adopting frequency of use as a primary outcome that is far more sensitive to the effects of treatment; however, we note that the clinical meaningfulness of frequency measures of use remains to be established. In the next section, we demonstrate and highlight the degree to which use of a quantitative measure of frequency of use can produce discernible increases in statistical power in drug use disorder trials.

2. The Cost of Dichotomizing

It is well known that dichotomization of a quantitative variable results in a loss of information. However, what is less widely recognized is the *degree* to which use of a quantitative measure of frequency of use, relative to a binary measure of abstinence, can yield very discernible gains in information and increases in statistical power for assessing efficacy of treatments. As mentioned previously, this is primarily driven by the relatively low rate of abstinence observed in many drug use disorder trials. From a statistical perspective, the loss of power from dichotomizing can be demonstrated by considering a continuous or quantitative random variable (denoting frequency of drug use) having a standard normal distribution. The relationship between a standardized difference (scaled by the standard deviation) between treatment groups on the continuous scale, denoted by δ , and the

corresponding standardized difference after dichotomizing at cut-point d , denoted by Δ^d , can be written as

$$\Delta^d = \sqrt{R} \Delta^c$$

with $R = h^2/p(1-p)$ and where p is the proportion (or prevalence) of individuals below the cut-point and h is the ordinate from the standard normal density at the cut-point d (Cox, 1957; MacCallum et al., 2002); see Table 1 for definitions of these parameters. It can be shown that $0 < R < .64$, and therefore the standardized difference in proportions is *always* less than the standardized difference in the means of the quantitative variable (MacCallum et al., 2002); consequently, the power of the two-group test based on the dichotomized variable is *always* lower than the two-group test based on the quantitative variable.

This result should not be surprising; however, what is less well appreciated is the *magnitude* of loss of information through dichotomizing when the prevalence of the dichotomy (e.g., abstinence) is relatively low. A simple related expression for the percentage of information lost due to dichotomizing is $(1-R) \times 100\%$; alternatively, R^{-1} can be interpreted as the factor by which the total sample size (n) must be inflated to offset the impact of dichotomizing. In this illustration (Figure 1) with a normal distribution, the loss is minimized when 50% are below the cut-point, $p = 0.5$ and $h = 0.399$, and the percentage of information lost is 36.3% (Figure 1(a)); this loss is akin to discarding data on over a third of the participants. Put another way (Figure 1(b), a total sample of $n=157$ participants ($R^{-1} = 1/0.637 = 1.57$) would be required to mitigate the impact of dichotomizing the quantitative variable in a sample of $n=100$ participants. All other cut-points lead to a *greater* loss of information. In many drug use disorder trials abstinence rates of 15% are quite typical (e.g., Carroll et al., (2014) report abstinence rates for 5 trials of cocaine use disorder that range from 5% to 20%, with mean $\approx 15\%$), yielding $p = 0.15$, $h = 0.232$, and the percentage of information lost is 57.8%; so, a sample of 237 participants would be required to mitigate the impact of dichotomizing in a sample of 100 participants. Figure 1(a) displays the loss of information as a function of the proportion above or below different cut-points for dichotomizing. Note that the loss increases rapidly as the proportion approaches 0 (or 1). Similarly, Figure 1(b) displays the factor by which sample size must be inflated to offset the impact of dichotomizing; note, the degree of inflation is approximately 60–85% for proportions greater than 0.25 (or less than 0.75) but then increases very rapidly as the proportion approaches 0 (or 1). Overall, these results indicate that use of a quantitative measure of frequency of use, rather than a binary measure of abstinence, can significantly reduce the required sample size in drug use disorder trials, often by a factor of more than a half, while retaining the same statistical power to detect treatment effects.

We note that this demonstration assumed a variable with a normal distribution; the assumption of normality here is arguably a convenient fiction. However, as we will show, the shape of the distribution matters little when the cut-point is made at an extreme. In substance use trials, frequency of use over the course of follow-up is a quantitative measure, anchored at the two extremes of “never” (or abstinent) and “always”. Next, we highlight the potential loss of information using data on cocaine use from an existing clinical trial, the National

Institute on Drug Abuse Collaborative Cocaine Treatment Study (CCTS). Full details on the design and procedures of the CCTS can be found in previous publications on the trial (e.g., Crits-Christoph et al., 1999). Briefly, the CCTS was a multi-site randomized clinical trial that compared the efficacy of four psychosocial treatments for cocaine dependence. Each treatment consisted of 6 months of active phase treatment. A total of 487 outpatients, who were diagnosed with DSM-IV cocaine dependence and had used cocaine during the past 30 days, were randomly assigned to one of the four treatment conditions. A composite cocaine use outcome measure, constructed by pooling information from self-report data and weekly observed urine samples, was used to code each month of treatment as abstinent versus any cocaine use.

Thus, in the CCTS there were 6 monthly assessments of cocaine use during follow-up yielding a quantitative measure of frequency of use with 7 distinct values, ranging from 0–6. Abstinence is defined when this quantitative measure is dichotomized in terms of 0 (complete abstinence during the 6 months) versus 1–6 (any use during the 6 months). We examined the proportion of participants in the CCTS with values of 0 to 6 months of use; they were 16.4%, 13.8%, 12.5%, 10.8%, 12.7%, 12.3%, and 21.4% respectively. Note that in the CCTS only slightly more than 15% of participants attained full abstinence during the 6 months of treatment. Based on these percentages, we can quantify the cost of dichotomizing rather than basing an analysis on a quantitative measure of frequency of use. Specifically, with only 7 discrete values, we can regard frequency of use as an ordinal variable and the relative loss of efficiency (or loss of information) from dichotomizing a 7-point ordinal scale at 0 is given by the following simple expression (see, for example, Whitehead, 1993):

$$R^{-1} = \frac{1 - \sum_{k=0}^6 \pi_k^3}{1 - \pi_0^3 - \left(\sum_{k=1}^6 \pi_k \right)^3}$$

where, π_k denotes the probability or proportion of participants with outcome in the k^{th} ordinal category ($k = 0, \dots, 6$). Specifically in this formula, π_0 denotes the probability of abstinence. For the estimated proportions obtained in the NIDA CCTS this yields a value of $R^{-1} = 2.38$. That is, a study with a dichotomous measure of abstinence requires almost 2½ times the total number of participants than a study with a 7-level ordinal measure of frequency of use. We note that this is entirely consistent with the loss of information highlighted in Figure 1(a) when dichotomization yields a prevalence below the cutoff of approximately 15% (recall, in CCTS, the prevalence of abstinence is 16.4%). This result, together with Figures 1(a) and 1(b), highlights the very extreme loss of information from reliance on abstinence as a primary outcome rather than a quantitative measure of frequency of use. In the CCTS the loss of information is exacerbated by the relatively low rate of full abstinence; a similar loss would arise from other dichotomizations that result in a single dominant category. For example, the prevalence of “any abstinence” (i.e., abstinent for at least 1 month) during the 6 months of treatment is 78.6% and yields a value of $R^{-1} = 1.93$. Thus, a study with a dichotomous measure of any abstinence requires almost twice the total number of participants than a study with a 7-level ordinal measure of frequency of use. In

contrast, the loss of information is less substantial when dichotomization results in categories that are more equally prevalent. For example, in the CCTS, the prevalence of “at least 3 months of abstinence” is 53.5% and yields $R^{-1} = 1.31$, corresponding to an approximate 30% increase in sample size.

3. Conclusion

While sustained abstinence is the outcome the FDA currently favors for drug use disorder trials, continued use of a dichotomous outcome, especially one made at the most extreme end of the scale for frequency of use, has considerable costs in terms of statistical power. The loss of information from reliance on this dichotomy is particularly acute in drug use disorder trials due to the relatively low prevalence of abstinence in these trials. Thus, while many readers will know that such dichotomization results in some loss of power, the full extent of the loss may not be fully appreciated. One important consequence of this loss of information is that clinical trials employing dichotomous outcomes require larger numbers of participants, often more than twice as many, creating greater challenges for recruitment and retention of trial participants. This suggests that statistical as well as subject-matter considerations should be incorporated into the choice of primary outcome for drug use disorder trials. Indeed, we see no contradiction in basing the primary analysis of a trial on a quantitative measures of frequency of use on statistical grounds, while also having interest in producing simple tables or summaries based on the threshold of abstinence to help communicate findings from the analyses that form the basis for inferences about treatment effects. Beyond statistical considerations of sample size and power, discerning the clinical meaningfulness of any endpoint, whether dichotomous or quantitative, requires gathering data that demonstrates its relationship to various longer-term follow-up indicators of response to treatment (Carroll et al., 2014; Crits-Christoph et al., 2013; Kiluk et al., 2016; Winchell et al., 2012), e.g., 12-month post-treatment follow-up measures of problems and functioning. For example, Carroll et al., (2014) evaluated the relationships between 15 candidate during-treatment endpoints, both continuous and dichotomous, and long term functional outcomes. Although these analyses did not yield a single (during-treatment) endpoint that was overwhelmingly superior to all others in predicting long term outcomes, more empirical work along these lines to determine the clinical significance of alternative (during-treatment) endpoints remains to be done.

Acknowledgments

Role of Funding Source

Funding for this study was provided by a National Institute of Health, National Institute on Drug Abuse grant (NIDA R21 DA042847, U10 DA015831, K24 DA022288). The National Institute of Health had no further role in study design, in the collection, analysis and interpretation of data, in the writing of the report, or in the decision to submit the paper for publication.

The authors thank Dr. Paul Crits-Christoph for sharing data from the NIDA Collaborative Cocaine Treatment Study. This paper is based in part on a presentation at the ACTTION/MOST (Measures of Outcome for Stimulant Trials) meeting at FDA White Oak Campus, Silver Spring, MD on October 18, 2016.

References

- Altman DG, Royston P. The cost of dichotomising continuous variables. *Brit Med J*. 2006; 332:1080. [PubMed: 16675816]
- Carroll KM, Kiluk BD, Nich C, DeVito EE, Decker S, LaPaglia D, Duffey D, Babuscio TA, Ball SA. Toward empirical identification of a clinically meaningful indicator of treatment outcome: Features of candidate indicators and evaluation of sensitivity to treatment effects and relationship to one year follow up cocaine use outcomes. *Drug Alcohol Depend*. 2014; 137:3–19. [PubMed: 24556275]
- Cohen J. The cost of dichotomization. *Appl Psychol Meas*. 1983; 7:249–253.
- Cox DR. Note on grouping. *J Am Stat Assoc*. 1957; 52:543–547.
- Crits-Christoph P, Gallop R, Gibbons MBC, Sadicario JS, Woody G. Measuring outcome in the treatment of cocaine dependence. *Alcohol Drug Depend*. 2013; 1:108.
- Crits-Christoph P, Siqueland L, Blaine J, Frank A, Luborsky L, Onken LS, Muenz LR, Thase ME, Weiss RD, Gastfriend DR, Woody GE, Barber JP, Butler SF, Daley D, Salloum I, Bishop S, Najavits LM, Lis J, Mercer D, Griffin ML, Moras K, Beck AT. Psychosocial treatments for cocaine dependence: National Institute on Drug Abuse collaborative cocaine treatment study. *Arch Gen Psychiatry*. 1999; 56:493–502. [PubMed: 10359461]
- Donovan DM, Bigelow GE, Brigham GS, Carroll KM, Cohen AJ, Gardin JG, Hamilton JA, Huestis MA, Hughes JR, Lindblad R, Marlatt GA, Preston KL, Selzer JA, Somoza EC, Wakim PG, Wells EA. Primary outcome indices in illicit drug dependence treatment research: Systematic approach to selection and measurement of drug use endpoints in clinical trials. *Addiction*. 2012; 107:694–708. [PubMed: 21781202]
- Fedorov V, Mannino F, Zhang R. Consequences of dichotomization. *Pharm Stat*. 2009; 8:50–61. [PubMed: 18389492]
- Food and Drug Administration, P.D.A.C. FDA Psychopharmacologic Drugs Advisory Committee: Buprenorphine (buprenorphine hydrochloride subdermal implant) for maintenance treatment of opioid dependence. Food and Drug Administration; Silver Spring, MD: 2013.
- Kiluk BD, Carroll KM, Duhig A, Falk DE, Kampman K, Lai S, Litten RZ, McCann DJ, Montoya ID, Preston KL, Skolnick P, Weisner C, Woody G, Chandler R, Detke MJ, Dunn K, Dworkin RH, Fertig J, Gewandter J, Moeller FG, Ramey T, Ryan M, Silverman K, Strain EC. Measures of outcome for stimulant trials: ACTION recommendations and research agenda. *Drug Alcohol Depend*. 2016; 158:1–7. [PubMed: 26652899]
- MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. *Psychol Methods*. 2002; 7:19. [PubMed: 11928888]
- Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: A bad idea. *Stat Med*. 2006; 25:127–141. [PubMed: 16217841]
- Streiner DL. Breaking up is hard to do: The heartbreak of dichotomizing continuous data. *Can J Psychiatry*. 2002; 47:262–266. [PubMed: 11987478]
- Tai, B., Chiang, CN., Bridge, P. Medication development for the treatment of cocaine dependence: issues in clinical efficacy trials. US Department of Health and Human Services, National Institutes of Health, National Institute on Drug Abuse, Office of Science Policy and Communications; 1997.
- Tiffany ST, Friedman L, Greenfield SF, Hasin DS, Jackson R. Beyond drug use: A systematic consideration of other outcomes in evaluations of treatments for substance use disorders. *Addiction*. 2012; 107:709–718. [PubMed: 21981638]
- Van den Brink W, Montgomery SA, Van Ree JM, van Zwieten-Boot BJ. On behalf of the Consensus Committee. ECNP consensus meeting March 2003 guidelines for the investigation of efficacy in substance use disorders. *Eur Neuropsychopharmacol*. 2006; 16:224–230. [PubMed: 16442270]
- Vocci, F., de Wit, H. Consensus statement on evaluation of outcome of pharmacotherapy for substance abuse/dependence: Report from a NIDA/CPDD meeting. National Institute on Drug Abuse Medications Development Division; Bethesda, MD: 1999.
- Whitehead J. Sample size calculations for ordered categorical data. *Stat Med*. 1993; 12:2257–2271. [PubMed: 8134732]
- Winchell C, Rappaport BA, Roca R, Rosebraugh CJ. Reanalysis of methamphetamine dependence treatment trial. *CNS Neurosci Ther*. 2012; 18:367–368. [PubMed: 22533722]

Highlights

- Abstinence is the gold standard in drug use disorder clinical trials
- However relatively few study participants ever achieve complete abstinence
- This has prompted clinical trialists to consider alternative endpoints
- We highlight the increase in statistical power with a measure of frequency of use
- Beyond statistical power, clinical meaningfulness of endpoints must be considered

Figure 1a

Figure 1b

Figure 1.

Figure 1a. Percentage of information lost when dichotomizing, based on various cut-points

Figure 1b. Sample size inflation factor to offset the impact of dichotomizing, based on various cut-point.

Table 1

Definitions of parameters used to characterize the information loss or sample size inflation due to dichotomizing a continuous (standard normal) outcome.

Parameter	Definition
c	Standardized difference between treatment groups on continuous outcome
d	Standardized difference between treatment groups on dichotomous outcome, based on cut-point d
d	Cut-point on continuous scale used to form dichotomy
p	Proportion of participants with values below cut-point d
h	Ordinate (or height) of standard normal density at cut-point d
$R = h^2/p(1-p)$	Relative efficiency (or comparative precision) of dichotomized analysis
$(1-R) \times 100\%$	Percentage of information lost due to dichotomizing
R^{-1}	Sample size inflation factor to offset impact of dichotomizing