HHS Public Access

Author manuscript

Am J Drug Alcohol Abuse. Author manuscript; available in PMC 2019 August 02.

Published in final edited form as:

Am J Drug Alcohol Abuse. 2019 ; 45(5): 451–459. doi:10.1080/00952990.2019.1584202.

Consumption outcomes in clinical trials of alcohol use disorder treatment: Consideration of standard drink misestimation

Megan Kirouac, PhD[a], **Eric Kruger, DPT**[a], **Adam D. Wilson, MS**[a], **Kevin A. Hallgren, PhD**[b], **Katie Witkiewitz, PhD**[a]

[a]Center on Alcoholism, Substance Abuse, and Addictions, University of New Mexico, Albuquerque, NM, USA;

[b]Department of Psychiatry and Behavioral Sciences, University of Washington, School of Medicine, Box 356560 NE Pacific St., Seattle, WA, USA

Abstract

Background.—The Food and Drug Administration recently added a new clinical endpoint for evaluating the efficacy of alcohol use disorder (AUD) treatment that is more inclusive of treatment goals besides abstinence: no heavy drinking days (NHDD). However, numerous critiques have been noted for such binary models of treatment outcome. Further, there is mounting evidence that participants inaccurately estimate the quantities of alcohol they consume during drinking episodes (i.e., drink size misestimation), which may be particularly problematic when using a binary criterion (NHDD) compared to a similar, continuous alternative outcome variable: percent heavy drinking days (PHDD). Yet, the impact of drinking misestimation on binary (e.g., NHDD) versus continuous outcome variables (e.g., PHDD) has not been studied.

Objectives.—Using simulation methods, the present study examined the potential impact of drink size misestimation on NHDD and PHDD.

Methods.—Data simulations were based on previously published findings of the amount of error in how much alcohol is actually poured when estimating standard drinks. We started with self-reported daily drinking data from COMBINE study participants with complete data ($N = 888$; 68.1% male), then simulated inaccuracy in those estimations based on literature on standard drink size misestimation.

Results.—Clinical trial effect sizes were consistently lower for NHDD than for PHDD. Drink size misestimation further lowered effect sizes for NHDD and PHDD.

Conclusions.—Drink size misestimation may lead to inaccurate conclusions about drinking outcomes and the comparative effectiveness of AUD treatments, including inflated type-II error rates, particularly when treatment “success” is defined by binary outcomes such as NHDD.

CONTACT Megan Kirouac mkirouac@unm.edu Center on Alcoholism, Substance Abuse, and Addictions, University of New Mexico, 2650 Yale Blvd SE, MSC 11-6280, Albuquerque, NM 87106, USA.

Disclosures

My co-authors and I do not have any conflicts of interest that could inappropriately influence, or be perceived to influence, our work.

Keywords

Alcohol use disorder; treatment outcomes; drink size; standard drink; statistical power; data simulation

Introduction

The Food and Drug Administration recently approved using the percentage of subjects with no heavy drinking days (NHDD) (i.e., no days with more than 3 or 4 standard drinks for women and men, respectively (1)), as a new endpoint for evaluating the efficacy of alcohol use disorder (AUD) medications. This approval marks an important shift away from abstinence as the only marker of success and towards accepting non-abstinence outcomes. However, there are numerous critiques of binary treatment outcomes, including the potential to create false dichotomies between "successes" and "failures" (2).

Dichotomizing continuous variables also has numerous statistical consequences (3–5), including the obscuring of individual differences, loss of reliability, reduced effect sizes, and loss of power. Researchers have cautioned against collapsing continuous drinking data (e.g., percentage of heavy drinking days; PHDD) into more coarse categories (6,7), noting potentially reduced effect sizes, which may be particularly detrimental for AUD clinical trials that often yield relatively small effect sizes (8,9). There is ample statistical evidence to conclude that dichotomizing continuous outcomes has a detrimental impact on effect size estimation; however, we do not know how much of a detriment this creates specifically when the continuous PHDD variable is dichotomized into NHDD.

The detrimental effects of collapsing continuous PHDD values into a binary NHDD outcome may be further impacted by participants' drink size misestimation. Drinking measures assume accurate "standard drink" size reporting by research participants – i.e., one "standard drink" containing exactly 14g of pure ethanol (http://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/standard-drink). However, participants have inconsistent conceptualizations of standard drink sizes, and typically inaccurately estimate their own drinking (10,11,12,13; see Table 1). There is mounting evidence that participants are unaware of standard drink definitions, that federal definitions of a "standard drink" are inconsistent with participants' own definitions of one drink, and that participants inaccurately estimate their drinking (10,11,12,13,14,15; see Table 1). These findings that participants misestimate their drink size have been consistent across cultures and across populations with varying levels of experience and training both consuming and pouring (e.g., bartenders) alcoholic beverages (10,11,12,13,14; see Table 1). Gender, drink type (e.g., wine, beer, spirits), and glass sizes are additional factors that further influence the amount of misestimation (13,14); however, across scenarios, participants typically underestimate (rather than overestimate) the number of drinks they consume.

With FDA approval for NHDD as a primary endpoint of AUD treatment trials, it is critical to understand how much drink size misestimation may impact clinical trial conclusions. Yet, no prior studies have examined the impact of drink size misestimation on NHDD or alternative non-abstinence endpoints. The current study compares the impact of drink size

misestimation on binary NHDD and continuous PHDD treatment outcomes using a simulation study that was informed by real data.

Methods

Participants

Self-reported daily drinking data were obtained from the first 90-day follow-up assessment in the COMBINE study ($N = 1383$; 16), including baseline and 90-day follow-up assessment data. Full sample data were examined for descriptive statistics. The COMBINE study was a multisite randomized clinical trial for participants meeting criteria for alcohol dependence (17). Treatments included combinations of medications (naltrexone, acamprosate, or placebo) and behavioral treatments (Medication Management or Combined Behavioral Intervention). Previous literature (16) found the largest changes in abstinence were for the naltrexone versus placebo sub-sample; therefore, these were the conditions compared when studying effect sizes in the present study. All treatments lasted 16-weeks; follow-up data were collected up to 12-months after treatment. See previous publications for description of IRB and ethics committee approval procedures (16).

Drinking data collection

The COMBINE study used a rigorous methodology for accurately collecting drinking data. Trained research assistants administered the Form 90, a calendar-based method that asks participants about their drinking in the 90 days preceding the assessment (18). Memory cues were used to facilitate accurate recollection and visual aids of drink containers were provided with active probing about drink sizes in effort to obtain accurate drink size estimates. Research assistants collected data on brands and types of beverages and computed the number of standard drinks rather than relying on participant calculations. The COMBINE study also used biochemical verification for participant drinking using % Carbohydrate Deficient Transferrin (%CDT; 16). For the purposes of our simulation, we considered the methodology used in the COMBINE study to be the gold standard method for accurately assessing drinking quantities in alcohol clinical trials. We therefore used a simulation design with the COMBINE data reflecting the “true” amount of alcohol consumption, which we then degraded by incorporating increasing levels of drink size misestimation.

Simulation design

Drink size misestimation parameters used in the simulation were based on research examining how much participants poured in drink containers versus how much they reported having poured (see Table 1; 11, 14, 19,20,21,22), which shows that participants usually underestimate the amount of alcohol poured. Because country, sample, and study methodology varied across studies that were reviewed (10,11,12,13,14; 18,19,20,21) and the amount of misestimation often varied by gender, glass size, and type of alcohol, we tested multiple degrees of drink size misestimation in the present simulation. The mean number of grams of alcohol poured (and average across-study standard deviations) from these studies were used to simulate distributions of drinking misestimation randomly sampled from gamma distributions to account for the positive skew that is typically observed in these

studies (for formula see: 23, p. 238). Studies were derived from a literature search for empirical research articles examining “standard drinks” and “drink size” estimation methodology. Values used in the present simulation models were derived from the overall average misestimation in each study, which was converted into a common metric (grams of ethyl alcohol) and calculated in terms of the proportion relative to the study’s country’s standard drink definition (e.g., 8g alcohol in the UK, 14g if the study was in the US). Six modal levels of drink size misestimation were simulated: 1.0, 1.2, 1.4. 1.6,1.8 and 2.0, which corresponded to standard drink size misestimation means of 1.24, 1.41, 1.59, 1.77, 1.95 and 2.14. These six modes were chosen to represent the variability of mean drink size misestimation represented in the literature (Table 1) that corresponds to simulating gamma distributions (which uses modes). A constant SD of 0.55 was used on all conditions to test how much of an effect varying levels of drink misestimation has on effect sizes while holding the amount of variability in misestimation between participants constant. Each of the six models had the following shape and rate parameters, respectively: 2.11, 4.11; 6.61, 4.67; 8.36, 5.26; 10.37, 5.85; 12.63, 6.46; 15.16, 7.08. For each condition of the simulation, 10,000 simulations were completed.

Drink misestimation coefficients were randomly drawn from a gamma distribution for each participant who had complete drinking data in the 90-day windows prior to baseline and post-treatment ($N = 888$; i.e., participants with missing data within the baseline or post-treatment assessment windows were excluded to minimize the impact of missing data on the simulation tested here). These drink misestimation coefficients reflected the number of standard drinks participants would be expected to pour in a free pour task based on previous research (reported in Table 1). For example, a mean coefficient of 1.25 would indicate that a participant poured 1.25 standard drinks when asked to pour 1 standard drink. Simulated misestimation of daily drinking was then computed by dividing the “true” number of standard drinks by the drink misestimation coefficient. For example, if a participant reported 10 standard drinks as their “true” alcohol consumption level in the COMBINE study, a drink misestimation coefficient of 1.25 would indicate that this participant would have *reported* consuming 8 standard drinks due to a drink size misestimation under less rigorous methodological conditions than the COMBINE study (e.g., simply asking a participant how many drinks they have consumed). By dividing the number of standard drinks in COMBINE by the drink misestimation coefficient, the present study simulates the underestimation of “true” standard drink consumption; stated differently, what ***would have been*** reported had the study used less rigorous methods of collecting drinking data given the tendency for participants to underestimate the number of standard drinks they consume.

PHDD and NHDD outcome variables were then derived for each simulated dataset; both PHDD and NHDD used gender-specific definitions for heavy drinking (24). Effect sizes comparing naltrexone versus placebo were computed for these simulated (misestimated) PHDD and NHDD values using Cohen’s d and Cohen’s h, for continuous and binary variables, respectively (25). We then averaged across simulations within each of the six conditions. Previous research has compared Cohen’s d values to those of Cohen’s h to examine the different effect sizes generated by using PHDD versus the percentage of

participants with NHDD (26). Simulations and analyses were conducted in R version 3.3 (27).

Results

Table 2 and Figures 1–2 show mean values of PHDD and NHDD for the original and simulated drinking data. Observed PHDD was 61.6% at baseline and 18.2% at post-treatment. Simulations with increasing misestimation in standard drink reporting reduced PHDD values to 52.6–30.3% at baseline and 15.5–7.0% at post-treatment. The original percentage of subjects with NHDD was 0.05% and 43.0% at baseline and post-treatment. Simulations with increasing misestimation in standard drink reporting increased these values to 5.3–20.7% at baseline and 50.2–71.0% at post-treatment.

Effect sizes comparing naltrexone versus placebo were directly impacted by drink size misestimation, with a greater impact observed for NHDD than for PHDD (see Table 2 and Figure 3). The effect size for PHDD in the original data was −0.088 and ranged from −0.089 to −0.056 across the simulated misestimation conditions (see Figure 3). The effect size for NHDD in the original data was −0.055 and ranged from −0.055 to −0.019 across the simulated misestimation conditions (see Figure 3). Effect sizes for PHDD decreased by as much as 36% when drink misestimation was greatest (simulation 6). Effect sizes for NHDD were more drastically impacted at each level of the simulation, with greatest decrease in effect size at maximum drink misestimation (simulation 6, 65% decreased effect size). At almost every change in drink misestimation simulated, effect sizes of NHDD were weakened at nearly double the severity of those for PHDD: 5% versus 13%, 11% versus 24%, 17% versus 35%, 26% versus 49%, and 36% versus 65%. The sole exception was for simulation 1, where drink misestimation was smallest, effect sizes were negligibly impacted for PHDD and NHDD (1% improvement and no change, respectively).

Discussion

Evaluating AUD treatment efficacy by dichotomizing self-reported drinking data involves at least two potential perils: loss of statistical power and drink size misestimation impacts on effect sizes. The consequences of distilling continuous data into binary data were examined in addition to the potential consequences of drink size misestimation. Effect sizes were consistently smaller for binary NHDD than continuous PHDD and drink size misestimation further decreased effect sizes. Results indicated that both NHDD and PHDD are impacted by drink size misestimation at rates proportionate to the degree of misestimation. Together, these conditions could produce the appearance that many more patients achieve the criterion of NHDD at the end of AUD treatment (e.g., up to 71%) compared to the actual number of patients who achieved NHDD (e.g., 43%). The practical conclusions drawn from PHDD effects were consistently a minority of participants' days consisted of heavy drinking (18% PHDD and 7% PHDD for observed COMBINE data and simulation 6 data, respectively), thus drinking misestimation produced less drastic differences in treatment success rates as compared to NHDD. The combined impact of dichotomizing drinking data that are misestimated highlights the potential threat for researcher conclusions about treatment efficacy.

There was also some difference in the relative impact of drink size misestimation on standardized effect size estimates between treatment conditions for NHDD compared to PHDD. For example, even a slight standard drink consumption misestimation of 1.41 drinks (simulation 2) detrimentally impacted the relative change in effect sizes for NHDD nearly three-times more than effect sizes of PHDD (effect sizes reduced by 5% for PHDD compared to 13% for NHDD compared to original COMBINE data). Although overall effect size values were small, these results demonstrate a potential for differential impact of drink size misestimation for binary versus continuous outcomes.

The mechanisms through which drink size misestimation and variable dichotomization may impact effect sizes may include several components manipulated through the present study's simulation methodology. First, participants generally underreport their alcohol consumption (reflected by the mean bias parameter) and variability in the degree of misestimation between participants (reflected by the bias standard deviation parameter) adds additional statistical "noise." Second, the dichotomization of continuous daily drinking data into binary heavy or non-heavy drinking days may introduce errors in which many heavy drinking days become misclassified as non-heavy days. The reduction of continuous PHDD into a single binary value of non-heavy or heavy drinking further reduces the amount of information available for analysis. Given the complexity of drinking data, distilling data into binary variables poses many potential pitfalls for decreased accuracy of data, especially when considering drink size misestimation. Although the lines graphed in Figure 3 portray the impacts of drink misestimation as appearing equitable for PHDD and NHDD, examination of the raw data highlight the differential impacts on the two outcomes, primarily due to the original loss of power with NHDD compared to PHDD. Since the original COMBINE data showed effect size of NHDD as −0.055, there was a relatively weaker margin of error compared to the original COMBINE effect size of PHDD of −0.088. A change of 0.02, therefore impacted NHDD proportionately greater than PHDD. Therefore, using continuous outcomes, such as PHDD, has more likelihood to preserve the integrity of complex drinking data and may explain why effect sizes for PHDD were somewhat less detrimentally impacted by drink size misestimation than those of NHDD.

Limitations

One limitation to the present findings is the assumption that the COMBINE data reflected accurate drinking quantities. However, this assumption was the most straightforward methodology and is supported by the rigorous data collection methodology with biochemical verification employed by the COMBINE study team (16). Another limitation is that effect sizes were small in COMBINE (16) and even smaller in the simulated data examined in the present study. The present findings merely highlight the proportionate impact of drink misestimation on the binary NHDD versus PHDD. Using another dataset that had larger treatment effects may have provided more meaningful evidence of the impact of drink misestimation. Further, the study samples in the drink misestimation literature were not clinical samples like that in COMBINE. It is theoretically possible that individuals with AUD diagnoses may be more or less accurate in their reporting of standard drinks; future research should examine how misestimation may vary as a function of stage of treatment or recovery status. Third, the studies included in Table 1 are not comprehensive of the entire

drink mis-estimation literature and studies that did not report drink misestimation standard deviations were omitted from the present study (e.g., 28,29,15). However, a recent systematic review suggests average misestimation values fall within similar ranges of those modeled in the present study (see 10 for systematic review). Moreover, the present methodology simulates varying levels of misestimation, which was designed to provide readers with extrapolatable information to determine what kinds of impacts to their data might be expected if drink size misestimations are outside of the exact simulated values in the present study. Another limitation was that we assumed within-participant estimation was consistent across and within drinking occasions within 90-day assessments. Future research may aim to quantify the extent to which within-participant drink size misestimation varies between each drink consumed, as has been done across drink types (e.g., wine, beer, spirits) and drink glass sizes (e.g., 14,15).

Conclusions and recommendations

The present study examined the impact of variable dichotomization and drink size misestimation on outcomes used to examine the efficacy of treatments for AUD: no heavy drinking days (NHDD) and percent heavy drinking days (PHDD). Results indicated that the effect sizes of naltrexone versus placebo for the binary outcome of NHDD were overall lower than those for PHDD. Drink size misestimation further decreased effect sizes. Specifically, increasing levels of drinking data misestimation decreased the treatment effect sizes for NHDD at approximately twice the impact of that compared to effect sizes for the continuous PHDD variable. Such findings provide caution for future research in considering both measurement methods where drink size misestimation may be more prevalent (e.g., quantity-frequency questionnaires (30–36) and outcome variable selection (e.g., variable dichotomization).

Based on the present findings and the need to use a consistent “yardstick” in reporting treatment outcomes for AUD treatment efficacy studies (37), the increased power of using continuous PHDD highlights the potential danger of using dichotomous NHDD as the sole determination of treatment efficacy since NHDD will inherently have less statistical power to demonstrate treatment effects. Moreover, non-addiction treatment efficacy studies (e.g., chronic obstructive pulmonary disease, weight management, pulmonary arterial hypertension, depression medication trials) do not regularly use binary endpoints in their research (38–41). That addiction research stands alone in healthcare research as using a binary outcome that inherently has less statistical power than alternative, continuous outcomes is a readily correctable limitation of our current science.

Perhaps a less easily correctable limitation of our current AUD treatment science is the limitation of using self-reported drinking data, which holds potential for drink size misestimation. Given that the efficacy of AUD treatments is often tested by examining alcohol consumption, future research may benefit from exploring new data collection methodologies, such as real-time monitoring of drinking behavior (42). At a policy level, printing standard drink information on alcohol container labels may improve drink reporting accuracy (43). Future research may also benefit from broadening the conceptualization of treatment outcomes to consistently include non-consumption outcomes, such as

psychosocial functioning and quality of life (44,45). Such non-consumption definitions would not only address the limitations inherent with drink size misestimation but would also address decades' worth of researchers' calls for more client individualized definitions of treatment success (44–46). Additional research in identifying measures of non-consumption outcomes with the best psychometric properties and greatest sensitivity and specificity for short- and long-term outcomes is needed so AUD treatment researchers can adopt consistent "yardsticks of success" (37; 47,48,49).

Acknowledgments

This research was supported by grants from the National Institute on Alcohol Abuse and Alcoholism (NIAAA): F31-AA024959, PI: Kirouac; K01-AA024796, PI: Hallgren; R01-AA022328 and R01-AA025539, PI: Witkiewitz; F31-AA 026773, PI: Wilson). Adam D. Wilson was supported by a training grant from NIAAA (T32-AA0018108; PI: McCrady). We would also like to acknowledge Matthew R. Pearson, Adrian J. Bravo, and Mark Prince for their statistical consultation for data simulation methodology used in the present manuscript.

Funding

This research was supported by grants from the National Institute on Alcohol Abuse and Alcoholism (NIAAA): F31-AA024959, PI: Kirouac; F31-AA026773, PI: Wilson; K01-AA024796, PI: Hallgren; R01-AA022328 and R01-AA025539, PI: Witkiewitz; Adam D. Wilson was also supported by a training grant from NIAAA (T32-AA0018108; PI: McCrady).

References

1. Guidance for Industry. Alcoholism: developing drugs for treatment Center for Drug Evaluation and Research (CDER). Rockville, MD: Food and Drug Administration; 2015 Feb.
2. Wilson AD, Bravo AJ, Pearson MR, Witkiewitz K. Finding success in failure: using latent profile analysis to examine heterogeneity in psychosocial functioning among heavy drinkers following treatment. Addiction. 2016 Dec 1;111(12):2145–54. doi:10.1111/add.13518. [PubMed: 27367263]
3. Altman DG, Royston P. The cost of dichotomising continuous variables. Bmj. 2006 May 4;332(7549):1080. doi:10.1136/bmj.332.7549.1080. [PubMed: 16675816]
4. Campbell MJ, Julious SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. Bmj. 1995 Oct 28;311(7013):1145–48. [PubMed: 7580713]
5. MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. Psychol Methods. 2002 Mar; 7(1):19. doi:10.1037/1082-989X.7.1.19. [PubMed: 11928888]
6. McGinley JS, Curran PJ. Validity concerns with multiplying ordinal items defined by binned counts. Methodology. 2014;10:108–16. doi:10.1027/1614-2241/a000081. [PubMed: 25383075]
7. McGinley JS, Curran PJ, Hedeker D. A novel modeling framework for ordinal data defined by collapsed counts. Stat Med. 2015 Jul 10;34(15):2312–24. doi:10.1002/sim.6495. [PubMed: 25857717]
8. Litten RZ, Castle IJ, Falk D, Ryan M, Fertig J, Chen CM, Yi HY. The placebo effect in clinical trials for alcohol dependence: an exploratory analysis of 51 naltrexone and acamprosate studies. Alcohol Clin Exp Res. 2013 Dec 1;37(12):2128–37. doi:10.1111/acer.12197. [PubMed: 23889231]
9. Fitzmaurice GM, Lipsitz SR, Weiss RD. Statistical considerations in the choice of endpoint for drug use disorder trials. Drug Alcohol Depend. 2017 Dec 1;181:219–22. doi:10.1016/j.drugalcdep.2017.09.031. [PubMed: 29102827]
10. Schultz NR, Kohn CS, Schmerbauch M, Correia CJ. A systematic review of the free-pour assessment: implications for research, assessment and intervention. Exp Clin Psychopharmacol. 2017 Jun; 25(3):125. doi:10.1037/pha0000120. [PubMed: 28287794]
11. Gual AN, Martos AR, Lligoña A, Llopis JJ. Does the concept of a standard drink apply to viticultural societies?. Alcohol Alcohol. 1999 Mar 1;34(2):153–60. [PubMed: 10344775]

12. Jones SC, Gregory P. The impact of more visible standard drink labelling on youth alcohol consumption: helping young people drink (ir) responsibly?. Drug Alcohol Rev. 2009 May 1;28(3): 230–34. [PubMed: 21462396]
13. Kerr WC, Stockwell T. Understanding standard drinks and drinking guidelines. Drug Alcohol Rev. 2012 Mar 1;31(2):200–05. doi:10.1111/j.1465-3362.2011.00374.x. [PubMed: 22050262]
14. Kerr WC, Greenfield TK, Tujague J, Brown SE. A drink is a drink? Variation in the amount of alcohol contained in beer, wine and spirits drinks in a US methodological sample. Alcohol Clin Exp Res. 2005 Nov 1;29(11):2015–21. doi:10.1097/01.alc.0000187596.92804.bd. [PubMed: 16340459]
15. Wansink B, Van Ittersum K. Shape of glass and amount of alcohol poured: comparative study of effect of practice and concentration. Bmj. 2005 Dec 22;331 (7531):1512–14. doi:10.1136/bmj.331.7531.1512. [PubMed: 16373735]
16. Anton RF, O'Malley SS, Ciraulo DA, Cisler RA, Couper D, Donovan DM, Gastfriend DR, Hosking JD, Johnson BA, LoCastro JS, et al. Combined pharmacotherapies and behavioral interventions for alcohol dependence: the COMBINE study: a randomized controlled trial. Jama. 2006 May 3;295(17):2003–17. doi:10.1001/jama.295.17.2003. [PubMed: 16670409]
17. American Psychiatric Association. Diagnostic and statistical manual of mental disorder, text revision (DSMIV-TR). Washington, DC: American Psychiatric Association; 2000 p. 739–41.
18. Miller WR Form 90: A structured assessment interview for drinking and related behaviors: test manual: US department of health and human services. Rockville, MD: Public Health Service, National Institutes of Health, National Institute on Alcohol Abuse and Alcoholism 1996.
19. Gill JS, Donaghy M. Variation in the alcohol content of a 'drink' of wine and spirit poured by a sample of the Scottish population. Health Educ Res. 2004 Oct 1;19(5):485–91. [PubMed: 15345708]
20. Gill J, O'may F. Practical demonstration of personal daily consumption limits: a useful intervention tool to promote responsible drinking among UK adults? Alcohol Alcohol. 2007 Jun 18;42(5):436–41. doi:10.1093/alcalc/agm049. [PubMed: 17576724]
21. Kerr WC, Patterson D, Koenen MA, Greenfield TK. Alcohol content variation of bar and restaurant drinks in Northern California. Alcohol Clin Exp Res. 2008 Sep 1;32(9):1623–29. doi:10.1111/acer.2008.32.issue-9. [PubMed: 18616674]
22. Wilkinson C, Allsop S, Chikritzhs T. Alcohol pouring practices among 65-to 74-year-olds in Western Australia. Drug Alcohol Rev. 2011 Mar 1;30(2):200–06. doi:10.1111/j.1465-3362.2010.00218.x. [PubMed: 21355907]
23. Kruschke J Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan. Academic Press; 2014 Nov 11.
24. US Department of Health and Human Services. National institute on alcohol abuse and alcoholism In10th special report to the US Congress on alcohol and health: highlights from current research. Rockville, MD; 2000 Jun.
25. Cohen J Statistical power analysis for the behavioral sciences. 2nd ed. New York, NY: Routledge; 2013 May 13.
26. Falk D, Wang XQ, Liu L, Fertig J, Mattson M, Ryan M, Johnson B, Stout R, Litten RZ. Percentage of subjects with no heavy drinking days: evaluation as an efficacy endpoint for alcohol clinical trials. Alcohol Clin Exp Res. 2010 Dec 1;34(12):2022–34. doi:10.1111/acer.2010.34.issue-12. [PubMed: 20659066]
27. Team RCore. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
28. Lemmens PH. The alcohol content of self-report and 'standard'drinks. Addiction. 1994 May;89(5): 593–601. [PubMed: 8044126]
29. White AM, Kraus CL, Flom JD, Kestenbaum LA, Mitchell JR, Shah K, Swartzwelder HS. College students lack knowledge of standard drink volumes: implications for definitions of risky drinking based on survey data. Alcohol Clin Exp Res. 2005 Apr; 29(4):631–38. doi:10.1097/01.ALC.0000158836.77407.E6. [PubMed: 15834229]

30. Collins RL, Parks GA, Marlatt GA. Social determinants of alcohol consumption: the effects of social interaction and model status on the self-administration of alcohol. J Consult Clin Psychol. 1985 Apr; 53(2):189. doi:10.1037/0022-006X.53.2.189. [PubMed: 3998247]
31. Greenfield TK. Ways of measuring drinking patterns and the difference they make: experience with graduated frequencies. J Subst Abuse. 2000 Sep 1;12(1–2):33–49. [PubMed: 11288473]
32. Greenfield TK, Nayak MB, Bond J, Ye Y, Midanik LT. Maximum Quantity Consumed and Alcohol-Related Problems: assessing the Most Alcohol Drunk With Two Measures. Alcohol Clin Exp Res. 2006 Sep 1;30(9):1576–82. doi:10.1111/acer.2006.30.issue-9. [PubMed: 16930220]
33. Rehm J, Greenfield TK, Walsh G, Xie X, Robson L, Single E. Assessment methods for alcohol consumption, prevalence of high risk drinking and harm: a sensitivity analysis. Int J Epidemiol. 1999 Apr 1;28(2):219–24. [PubMed: 10342682]
34. Romelsjö A, Leifman H, Nyström S. A comparative study of two methods for the measurement of alcohol consumption in the general population. Int J Epidemiol. 1995 Oct 1;24(5):929–36. [PubMed: 8557449]
35. Room R Measuring drinking patterns: the experience of the last half century. J Subst Abuse. 2000 Sep 1;12 (1–2):23–31. [PubMed: 11288472]
36. Utpala-Kumar RA, Deane FP. Rates of alcohol consumption and risk status among Australian university students vary by assessment questions. Drug Alcohol Rev. 2010 Jan 1;29(1):28–34. doi: 10.1111/j.1465-3362.2009.00082.x. [PubMed: 20078679]
37. Sobell LC, Sobell MB, Connors GJ, Agrawal S. Assessing drinking outcomes in alcohol treatment efficacy studies: selecting a yardstick of success. Alcohol Clin Exp Res. 2003 Oct; 27(10):1661–66. doi:10.1097/01.ALC.0000091227.26627.75. [PubMed: 14574238]
38. States United. Guidance for industry: chronic obstructive pulmonary disease: developing drugs for treatment. Rockville, MD: Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research; 2007. doi:10.1094/PDIS-91-4-0467B.
39. United States. Guidance for Industry: developing products for weight management. Rockville, MD: Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research; 2007. doi:10.1094/PDIS-91-4-0467B.
40. McGlinchey N, Peacock AJ. Endpoints in PAH clinical trials in the era of combination therapy: how do we decide whether something is working without going bankrupt? Drug Discov Today. 2014 Aug 1;19(8):1236–40. doi:10.1016/j.drudis.2014.04.020. [PubMed: 24814434]
41. United States. Major depressive disorder: developing drugs for treatment: guidance for industry. Rockville, MD: Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research; 2018.
42. Barnett NP, Celio MA, Tidey JW, Murphy JG, Colby SM, Swift RM. A preliminary randomized controlled trial of contingency management for alcohol use reduction using a transdermal alcohol sensor. Addiction. 2017 Jun 1;112(6):1025–35. doi:10.1111/add.13767. [PubMed: 28107772]
43. Hobin E, Vallance K, Zuo F, Stockwell T, Rosella L, Simniceanu A, White C, Hammond D. Testing the efficacy of alcohol labels with standard drink information and national drinking guidelines on consumers' ability to estimate alcohol consumption. Alcohol Alcohol. 2017 Aug 9;53(1):3–11. doi:10.1093/alcalc/agx052.
44. Kaskutas LA, Borkman TJ, Laudet A, Ritter LA, Witbrodt J, Subbaraman MS, Stunz A, Bond J. Elements that define recovery: the experiential perspective. J Stud Alcohol Drugs. 2014 Nov; 75(6):999–1010. [PubMed: 25343658]
45. Neale J, Finch E, Marsden J, Mitcheson L, Rose D, Strang J, Tompkins C, Wheeler C, Wykes T. How should we measure addiction recovery? Analysis of service provider perspectives using online Delphi groups. Drugs (Abingdon Engl). 2014 Aug 1;21 (4):310–23.
46. Moos RH, Finney JW. The expanding scope of alcoholism treatment evaluation. Am Psychol. 1983 Oct; 38(10):1036. doi:10.1037/0003-066X.38.10.1036. [PubMed: 6314859]
47. Kirouac M, Stein ER, Pearson MR, Witkiewitz K. Viability of the World Health Organization quality of life measure to assess changes in quality of life following treatment for alcohol use disorder. Qual Life Res. 2017 Nov 1;26(11):2987–97. doi:10.1007/s11136-017-1631-4. [PubMed: 28647889]

48. Kirouac M, Witkiewitz K. Revisiting the drinker inventory of consequences: an extensive evaluation of psychometric properties in two alcohol clinical trials. Psychol Addict Behav. 2018 Feb; 32(1):52. doi:10.1037/adb0000344. [PubMed: 29419311]

49. Del Boca FK, Darkes J. 'Nothing is more practical than a good theory': outcome measures in addictions treatment research. Addiction. 2012 Apr; 107(4):719–20. doi:10.1111/j.1360-0443.2011.03647.x. [PubMed: 22372698]

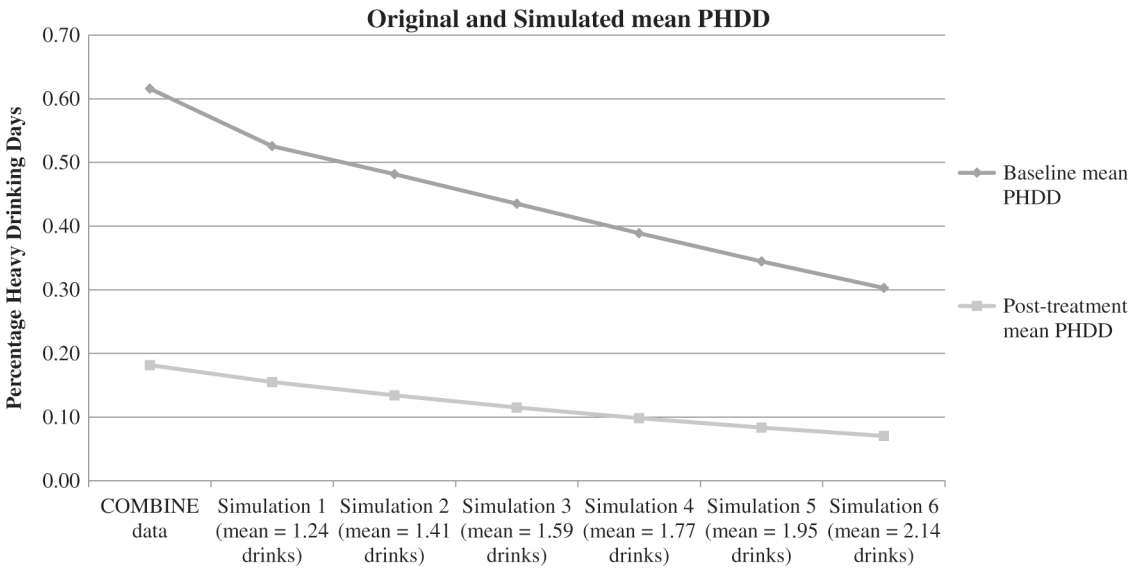

Figure 1.

Impact of drink size misestimation at baseline and post-treatment timepoints for PHDD.

Note: PHDD = percent heavy drinking days. Initial data point is the original COMBINE value; subsequent data points are for simulations 1 through 6, respectively.

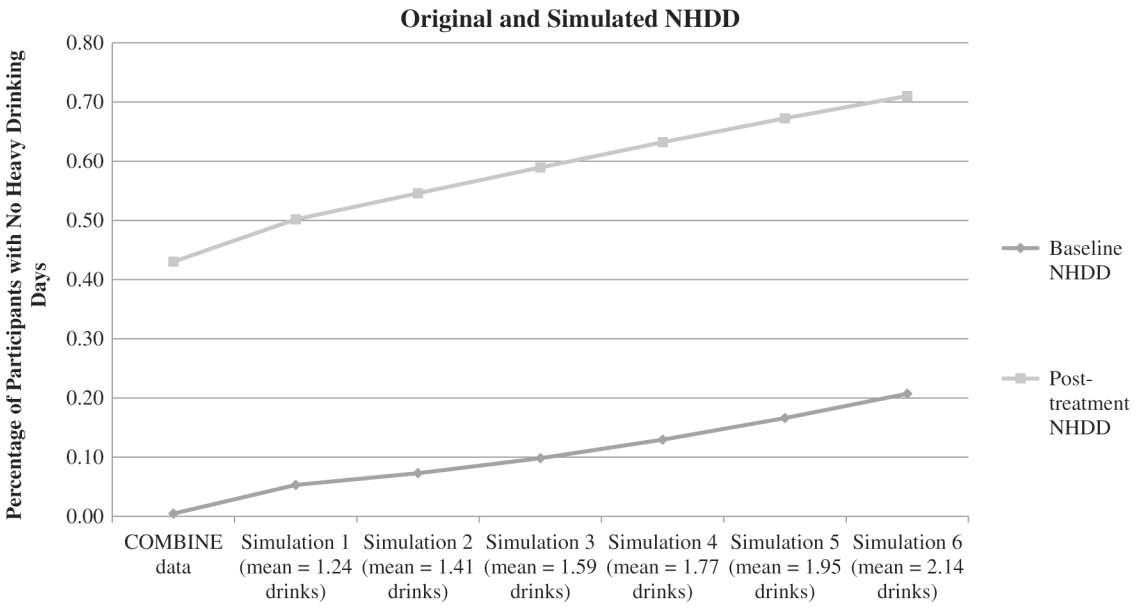

Figure 2.

Impact of drink size misestimation at baseline and post-treatment timepoints for NHDD.

Note: NHDD = percent subjects with no heavy drinking days. Initial data point is the original COMBINE value; subsequent data points are for simulations 1 through 6, respectively.

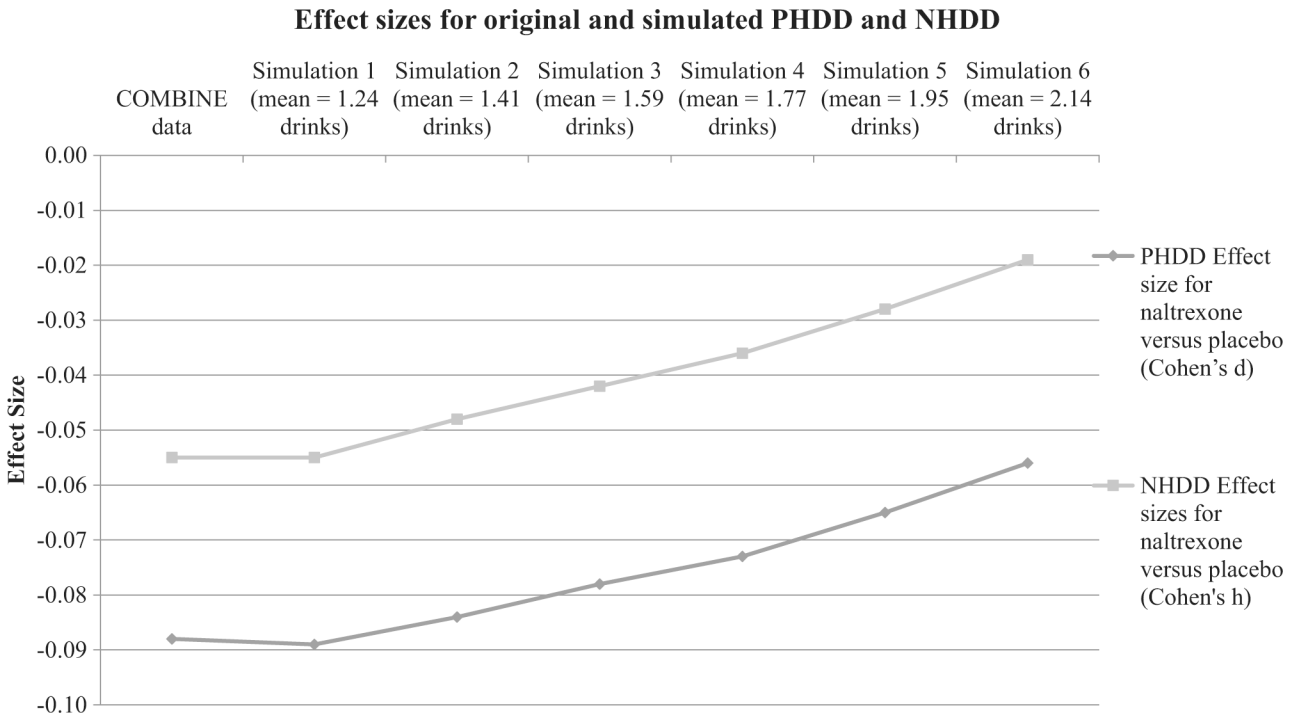

Figure 3.

Impact of drink size misestimation on naltrexone versus placebo effect sizes: PHDD and NHDD.

Note: PHDD = percent heavy drinking days: NHDD = percent subjects with no heavy drinking days. Initial data point is the original COMBINE value; subsequent data points are for simulations 1 through 6, respectively.

Table 1.

Previously published studies of the amount of alcohol poured into drink containers.

Citation	Sample	Methodology	Average number of standard drinks poured*	Standard deviation of number of drinks poured*
Gual et al., 1999	Field studies in Spain	Participants selected their 1 preferred glass type (of 16 commonly used glasses provided by researchers) and were asked to fill with the amount of alcohol they would typically consume at bars or at home.	1.375	0.541
Kerr et al., 2005	US national survey	Participants poured their “usual drink” into a beaker.	1.357	0.713
Gill & Donaghy., 2004	Regular drinkers in Scotland	Participants poured their “usual drink” into a glass.	1.832	0.545
Gill & O’May, 2007	Regular drinkers in the UK	Participants poured their “usual” amount of wine or spirits into a glass.	2.050	0.678
Kerr et al., 2008	Field studies in the US	Researchers went to 80 alcohol-serving establishments, ordered various drinks, and measured the bartenders’ pour.	1.615	0.040
Wilkinson et al., 2011	Current drinkers in Australia	Participants poured their “usual” serving of alcohol into their “usual drinking vessel” and then researchers measured the amount of alcohol poured.	1.221	0.466

Note:

*standard drink definitions were used based on the country in which the study was conducted

Am J Drug Alcohol Abuse. Author manuscript; available in PMC 2019 August 02.

Table 2.

Descriptive statistics for original COMBINE drinking data and simulated drinking data.

	Baseline mean PHDD	Post-treatment mean PHDD	PHDD Effect size for naltrexone versus placebo (Cohen's d)	Percent decrease in Cohen's d from COMBINE data	Baseline NHDD	Post-treatment NHDD	NHDD Effect size for naltrexone versus placebo (Cohen's h)	Percent decrease in Cohen's h from COMBINE data
COMBINE data	0.616	0.182	−0.088	–	0.005	0.430	−0.055	–
Simulation 1 data: mean = 1.24 drinks	0.526	0.155	−0.089	−1%	0.053	0.502	−0.055	0%
Simulation 2 data: mean = 1.41 drinks	0.482	0.134	−0.084	5%	0.073	0.546	−0.048	13%
Simulation 3 data: mean = 1.59 drinks	0.435	0.115	−0.078	11%	0.098	0.589	−0.042	24%
Simulation 4 data: mean = 1.77 drinks	0.389	0.098	−0.073	17%	0.129	0.632	−0.036	35%
Simulation 5 data: mean = 1.95 drinks	0.344	0.084	−0.065	26%	0.166	0.672	−0.028	49%
Simulation 6 data: mean = 2.14 drinks	0.303	0.070	−0.056	36%	0.207	0.710	−0.019	65%

Note: PHDD = percent heavy drinking days; NHDD = percent subjects with no heavy drinking days

Am J Drug Alcohol Abuse. Author manuscript; available in PMC 2019 August 02.