Published in final edited form as:

Alcohol Clin Exp Res. 2015 September ; 39(9): 1557–1570. doi:10.1111/acer.12800.

Recommendations for the Design and Analysis of Treatment Trials for Alcohol Use Disorders

Katie Witkiewitz, PhD¹, John W. Finney², Alex H.S Harris³, Daniel R. Kivlahan⁴, and Henry R. Kranzler⁵

¹Department of Psychology and Center on Alcoholism, Substance Abuse, and Addictions, University of New Mexico

²Center for Innovation to Implementation, VA Palo Alto Health Care System, Menlo Park, CA

³VA Substance Use Disorder Quality Enhancement Research Initiative, VA Palo Alto Health Care System, Menlo Park, CA

⁴Veterans Health Administration, Washington, DC and Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle

⁵Center for Studies of Addiction, Department of Psychiatry, University of Pennsylvania Perelman School of Medicine and VISN4 MIRECC, Philadelphia VAMC, Philadelphia, PA

Abstract

Background—Over the past 60 years the view that “alcoholism” is a disease for which the only acceptable goal of treatment is abstinence has given way to the recognition that alcohol use disorders (AUDs) occur on a continuum of severity, for which a variety of treatment options are appropriate. However, because the available treatments for AUDs are not effective for everyone, more research is needed to develop novel and more efficacious treatments to address the range of AUD severity in diverse populations. Here we offer recommendations for the design and analysis of alcohol treatment trials, with a specific focus on the careful conduct of randomized clinical trials of medications and non-pharmacological interventions for AUDs.

Methods—Narrative review of the quality of published clinical trials and recommendations for the optimal design and analysis of treatment trials for AUDs.

Results—Despite considerable improvements in the design of alcohol clinical trials over the past two decades, many studies of AUD treatments have used faulty design features and statistical methods that are known to produce biased estimates of treatment efficacy.

Conclusions—The published statistical and methodological literatures provide clear guidance on methods to improve clinical trial design and analysis. Consistent use of state-of-the-art design

Correspondence should be addressed to: Katie Witkiewitz, Department of Psychology, University of New Mexico, 1 University of New Mexico, MSC03 2220, Albuquerque, NM 87131-0001, Phone: 505-277-5953, Fax: 505-925-1394, katiw@unm.edu.

Disclosures: H. Kranzler has served as a consultant or advisory board member for Alkermes, Lilly, Lundbeck, Otsuka, Pfizer, and Roche. K. Witkiewitz has served as a consultant for Alkermes. H. Kranzler and K. Witkiewitz have received financial support from the American Society of Clinical Psychopharmacology’s Alcohol Clinical Trials Initiative (ACTIVE), which is supported by AbbVie, Ethypharm, Lilly, Lundbeck, and Pfizer. Some of the views on clinical trial design and methods reflect discussions that H. Kranzler and K. Witkiewitz had during their participation in ACTIVE meetings.

features and analytic approaches will enhance the internal and external validity of treatment trials for AUDs across the spectrum of severity. The ultimate result of this attention to methodological rigor is that better treatment options will be identified for patients with an AUD.

Keywords

alcohol randomized trials; research design; alcohol use disorder; eligibility criteria; data analysis; missing data approaches

The public health and societal costs of alcohol use disorders (AUDs) are considerable (U.S. Burden of Disease Collaborators, 2013). Alcohol use is the third leading global contributor to the burden of disease as measured in disability-adjusted life years (World Health Organization, 2009, 2011). The estimated cost of excessive alcohol consumption in the United States was \$223.5 billion in 2006 dollars (Bouchery et al., 2011). Yet, despite social and economic costs of heavy drinking, many individuals with AUDs never receive treatment (Substance Abuse and Mental Health Services Administration, 2014). For individuals with an AUD who seek care, relatively few efficacious treatments are available (Jonas et al., 2014a; National Institute for Health and Care Excellence, 2011; Zindel and Kranzler, 2014).

Currently, only three AUD medications are approved by the Food and Drug Administration (FDA): disulfiram, acamprosate, and naltrexone (including oral and long-acting injectable formulations), and one additional AUD medication, nalmefene, is approved by the European Medicines Agency (Mann et al., 2013). For a handful of other drugs that are not approved to treat AUDs (e.g., gabapentin, ondansetron, varenicline, topiramate), there is a varying amount of evidence of efficacy (Blodgett et al., 2014; Jonas et al., 2014; Zindel and Kranzler, 2014). The need for additional development of medications has been voiced by a collaboration of researchers, and representatives of the pharmaceutical industry, the FDA, and the National Institute on Alcohol Abuse and Alcoholism (NIAAA), who have collectively formed the Alcohol Clinical Trials Initiative (ACTIVE; Anton et al., 2012). Non-pharmacological treatments for AUD are also limited, with good evidence of efficacy for fewer than a dozen behavioral treatments (Miller and Wilbourne, 2002; National Institute for Health and Care Excellence, 2011). Most of the efficacious treatments yield small-to-moderate effect sizes, which may contribute to the fact that, in general, initiation and engagement in any of these treatment options is suboptimal (Brorson et al., 2013; Harris et al., 2012, 2009a). The low rates of utilization of treatments for AUDs and the modest efficacy of existing interventions underscore the need to develop and test new treatments that are more efficacious and attractive to patients and clinicians, including those that take into account the heterogeneity of AUDs (Anton et al., 2012; Lane and Sher, in press; Litten et al., 2015).

To enhance the validity of future AUD treatment trials, we review key elements of research design and analysis in both pharmacological and non-pharmacological alcohol treatment studies. Unlike prior reviews that covered design and analysis considerations for a wide variety of clinical conditions (Berger, 2006; Chu et al., 2011; Freemantle, 2001; Harris et al., 2009b) or that focused on limited design questions in alcohol treatment studies (e.g., intent-to-treat approaches, Del Re et al., 2013; randomization, Hedden et al., 2006), the current

review provides a broad review of recommendations that are comprehensive, but specific to alcohol treatment studies. We focus on four areas: (1) considerations for recruitment, randomization, and retention, (2) measurement selection for outcome assessment, intervention fidelity, and adherence monitoring, (3) timing of assessments and considerations for studying mechanisms of behavior change, and (4) statistical methods used to assess treatment efficacy.

Improvements in these areas are critical to advancing the methodological quality of alcohol clinical trials and the assessment of the reliability, validity, and scope of inference of treatment effects. Also, no matter how well studies are designed and conducted, it is critical to improve the quality of reporting of findings from AUD treatment trials. Witkiewitz and colleagues (this issue) make recommendations for improving the reporting of study features in four areas of relevance to alcohol treatment studies: (1) trial registration, (2) procedures for recruitment and retention, (3) procedures for randomization and intervention design, and (4) statistical methods used to assess treatment efficacy.

Considerations for Inclusion/Exclusion, Randomization, and Retention

A discussion of all relevant procedures for recruitment and data collection is beyond the scope of the current review. We focus here on three areas that have often been problematic in alcohol clinical trials: eligibility criteria, randomization procedures, and treatment retention. For additional coverage of this area, we recommend a prior review of methods for recruitment and retention in complex alcohol clinical trials (Zweben et al., 2005).

Eligibility criteria

The selection of participants for alcohol clinical trials is often biased by the exclusion of individuals with residential instability, low motivation, and co-occurring psychiatric, medical, or drug use disorders (Hoertel et al., 2014; Humphreys et al., 2008; Humphreys and Weisner, 2000; Rychtarik et al., 1998). Although exclusion criteria are often used in clinical trials to reduce within-group variance and thereby improve statistical power to detect clinically meaningful effects between groups (Food and Drug Administration, 1998), the application of exclusion criteria may also unintentionally impact the estimation of treatment effects intended to generalize to a less restricted population of treatment-seeking patients (Humphreys et al., 2008). The application of extensive exclusion criteria can reduce the external validity of clinical trial findings and contribute to the gap between clinical research and clinical practice (Kazdin, 2008; Newnham and Page, 2010).

Humphreys and colleagues (2008) found that commonly used eligibility criteria substantially influenced estimates of treatment effects in 25 treatment programs across the United States. The direction (positive or negative) and degree (small to large) of influence differed across two samples, suggesting that the impact of eligibility criteria on treatment outcomes cannot be reliably predicted across patient populations. We recommend that investigators consider the impact of exclusion criteria and document the reasons for excluding research participants (see Witkiewitz et al., this issue). Recommendations for the design of pragmatic trials, which emphasize real-world applications to inform intervention

delivery, could also be consulted to increase applicability of clinical trial findings to real-world settings (Loudon et al., 2015).

Randomization

Although the randomized clinical trial (RCT) is the “gold standard” design to evaluate treatment efficacy, it is not fail-safe. Thus, investigators need carefully to consider the choice of randomization procedures and assess whether randomization was successful. Hedden and colleagues (2006) reviewed several randomization options available for treatment trials for substance use disorders, the features of which are summarized in Table 1. The appropriate selection of a particular randomization scheme often depends on the size of the sample, potential imbalances that might occur across treatment and control groups, and *a priori* ideas about other variables (e.g., demographics, genotype) that might impact the effectiveness of the intervention (Kranzler et al., 2011).

The stratified permuted block randomization procedure is the most commonly used and recommended approach to adjust for covariate imbalances;; however, it is critical that the covariates used in stratification are included in the analysis and considered when conducting power analyses (Matts and Lachin, 1988). The use of permuted block and other covariate-adjusted randomization schemes can help to avoid differences in baseline characteristics across intervention groups, which can occur in the context of randomization failures that occur either systematically or by chance (Berger, 2005a). When baseline differences across groups are observed after randomization, propensity score or instrumental variable methods may be useful to estimate the degree of influence that the baseline characteristics have on treatment outcomes and to potentially reduce bias in treatment effect estimates (Berger, 2005b; Leyrat et al., 2013; Merkx et al., 2014). Propensity score or instrumental variable methods also can be useful when randomization is not feasible (Humphreys et al., 2014; Magura et al., 2013; Ye and Kaskutas, 2009). Methodological development in the application of instrumental variables and propensity score approaches is an area of active ongoing research (Austin, 2013; McCaffrey et al., 2013; Myers et al., 2011). As with any statistical technique, these methods can introduce bias if implemented incorrectly or if the instrumental variables are not well selected (Austin, 2008; Myers et al., 2011). Thus, thus investigators are encouraged to consult the most recent statistical literature when applying these methods. Sensitivity analyses, in which the outcomes are examined with and without the inclusion of covariates, and with and without the use of instrumental variables and/or propensity scores, can be used to gain a better understanding of the influence of these approaches on treatment effects.

Study retention and procedures to minimize missing data

It is often difficult to retain participants in alcohol clinical trials (Kranzler et al., 1996). Despite numerous procedures for handling missing data, described below, it is far preferable to prevent the occurrence of missing data by devoting resources to ensuring study retention. Recommendations to improve retention in medication trials include flexible dosing schedules, the continued assessment of individuals who discontinue treatment, the addition of the study treatment to an existing effective treatment, and the use of rescue medications or psychotherapy for individuals who do not respond to treatment (National Research

Council, 2010; Zweben et al., 2005). However, it is important to consider the unintended effects of these strategies on generalizability to real world patients and procedures (e.g., if the same strategies are not available in real world settings). Additional recommendations for reducing missing data include allowing follow-ups to be conducted via the phone or Internet (instead of only in-person), providing incentives/bonuses for completing assessments, using interim contacts with participants to maintain accurate contact information, asking participants to nominate “locators” who will know how to reach them should routine efforts prove unsuccessful, having and communicating to participants a clear project identity, and emphasizing to participants and staff the importance of engagement with the treatment (e.g., medication adherence), retention in the research, and the significance of the study (Davis et al., 2002; National Research Council, 2010). To gauge treatment effects accurately in RCTs, it is essential for investigators to attempt to follow-up all individuals at every time point regardless of whether they discontinued the intervention or did not receive the allocated intervention (National Research Council, 2010). Adaptive treatment protocols, such as those implemented in sequential multiple assignment randomized trial (SMART) designs, may also improve retention and generalizability by providing an individualized approach and by increasing treatment intensity for participants who need a higher level of care (Hinshaw et al., 2004; Lei et al., 2012; McKay, 2009; Murphy et al., 2007). Recommendations for the analysis of SMART designs should be carefully considered when such designs are employed (Chakraborty and Murphy, 2014; Nahum-Shani et al., 2012).

Some of the efforts to reduce attrition or to increase follow-up with individuals who have not received or completed treatment may influence outcomes. Thus, if efforts are implemented differentially across treatment groups (e.g., devoting more effort to tracking active intervention participants), then the estimation of treatment effects may also be impacted (Clifford et al., 2007; Kaminer et al., 2008; Maisto et al., 2007). Adequate blinding of research staff, particularly those involved with retention and assessment procedures, can greatly reduce the impact of differential retention efforts on the estimation of treatment effects. Nonetheless, the potential impact and burden of assessment procedures should be considered when designing the study (Gastfriend et al., 2005) and when evaluating the efficacy of the treatment and comparator (including placebo) conditions (Litten et al., 2013; Weiss et al., 2008). Secondary analysis of treatment effects with the number of assessments completed as a covariate could provide an indication of the degree to which the amount of assessment enhanced or attenuated treatment effects (Clifford and Davis, 2012; Kurtz et al., 2013).

Measurement Selection

Outcomes assessment

The most common primary outcome measures used in alcohol clinical trials are the frequency and/or intensity of alcohol consumption (Finney et al., 2003; Robinson et al., 2014), most often estimated using the Timeline Follow-back method (TLFB; Sobell and Sobell, 1992) or the Form-90 (Miller, 1996). These instruments provide daily alcohol consumption data, which are often aggregated to create the primary outcome measure. Commonly used aggregations include both continuous and binary outcomes. Continuous

outcomes include the percentage of days abstinent (PDA), drinks per drinking day (DDD), drinks per day (DPD), and the percentage of heavy drinking days (PHDD; with “heavy” defined as 4 or more drinks in a day for women and 5 or more drinks for men) (Anton and Randall, 2005). Binary outcomes include any drinking and no heavy drinking (Falk et al., 2010) and composite clinical outcomes (Cisler and Zweben, 1999). It is important to note that continuous outcomes allow for more powerful statistical tests and will often require smaller samples to detect treatment effects than binary ones (Bakhshi et al., 2012; MacCallum et al., 2002; Yoo, 2010). The selection of the primary outcome for a clinical trial may often depend on the focus of the trial (e.g., the PDA outcome may be more appropriate for abstinence-oriented treatments, while PHDD may be preferable for a moderation-focused treatment). With regard to instrument selection, we highly recommend that investigators collect drinking data on a daily level using a calendar method, such that various aggregations and analytic methods can be considered, depending on the goals of the trial (Hallgren et al., under review).

An important consideration in outcome selection is the identification of measures with established psychometric properties. Numerous studies have examined the reliability and validity of the TLFB and Form-90 across multiple administration methods, including comparisons between telephone and computer (Sobell et al., 1996), online and in-person (Pedersen et al., 2012), individual and group (Pedersen and LaBrie, 2006), and telephone versus self-administration (Maisto et al., 2008). The general conclusion from these studies is that both assessments, regardless of the method of administration, typically produce reliable and valid estimates of alcohol consumption in the context of clinical research (Del Boca and Darkes, 2003).

Self-reported alcohol consumption can be verified using collateral informants (i.e., proxy reporters) or alcohol biomarkers, though both approaches have limitations and currently most researchers ultimately rely on self-reported alcohol consumption for primary analyses. In general, collateral reports, such as family members, roommates or close friends, correspond moderately with self-reported alcohol use and related problems (Donovan et al., 2004; Whitford et al., 2009), but they may be skewed in the direction of less reporting (e.g., less alcohol use, higher functioning) than self-reports (Connors and Maisto, 2003). When collateral informants’ reports are used to validate self-reported drinking (Connors and Maisto, 2003; Donohue et al., 2004), it is important that the informants be confident in their knowledge of the patient’s drinking behavior (Sobell et al., 1997). Collateral reports have also been shown to be less helpful for confirmation of self-report in some populations (i.e., college students, Laforge et al., 2005).

Biomarkers of alcohol use are also commonly used to validate self-report data and the correspondence between the two is often quite high (Anton et al., 2006; Litten et al., 2010; UKATT Research Team, 2005). The most commonly used biomarkers, carbohydrate-deficient transferrin (CDT), gamma-glutamyltransferase, and the ratio of alanine aminotransferase to aspartate aminotransferase, are relatively inexpensive, but may be less sensitive to the level of alcohol consumed. Liver enzyme tests also have lower specificity than newer tests (Hashimoto et al., 2013; Hock et al., 2005), such as CDT and those measuring direct ethanol metabolites, including ethyl glucuronide, ethyl sulfate, and

phosphatidylethanol (Hashimoto et al., 2013; Jatlow et al., 2014). The newer biomarkers show good correspondence with self-reported alcohol consumption (Crunelle et al., 2014; Hartmann et al., 2007) and may be very useful to confirm recent self-reported abstinence, though they have limited sensitivity to detect the level of alcohol consumption or to confirm non-recent drinking via self-report (Jatlow et al., 2014), reducing their utility as primary outcome measures for alcohol clinical trials. Transdermal alcohol monitors are a promising new approach to collect real-time drinking data without reliance on self-report and have been used successfully to detect alcohol use in contingency management trials (Barnett et al., 2011; Dougherty et al., 2014). However, the available sensors may not be accurate in detecting lower levels of drinking (Barnett et al., 2014), are costly, and may not be acceptable to patients.

A range of secondary outcome measures also may be of interest, including alcohol-related problems, quality of life, psychosocial functioning, health care utilization, dependence severity, etc. (Cisler et al., 2005; LoCastro et al., 2009; Tiffany et al., 2012a). However, there is very little agreement regarding the “best” measures for assessing primary or secondary outcomes (Donovan, 2012; Tiffany et al., 2012b). The most common secondary outcome measures used in alcohol clinical trials include the Drinker Inventory of Consequences (DrInC; Miller et al., 1995); the Short Index of Problems, a shortened version of the DrInC (SIP; Feinn et al., 2003; Kiluk et al., 2013); the Addiction Severity Index (ASI; McLellan et al., 1992); measures of craving (e.g., the Obsessive Compulsive Drinking Scale, Anton, 2000; Impaired Control Scale, Heather et al., 1993) and other measures of functioning (e.g., Short Form Health Surveys, Ruta et al., 1993; Ware et al., 1996) or quality of life (e.g., Brief Symptom Inventory, Derogatis, 1983; World Health Organization Quality of Life scale, World Health Organization, 1998). Evidence supports the reliability and validity of these measures (Cacciola et al., 2011, 2007; Forcehimes et al., 2007), the utility of including non-consumption outcomes in large-scale alcohol clinical trials (Cisler et al., 2005; LoCastro et al., 2009), and their correspondence with alcohol consumption measures (Donovan et al., 2006; Witkiewitz, 2013).

Measures of intervention fidelity and medication adherence

During the course of an intervention it is important to monitor the fidelity of its delivery. In non-pharmacological interventions, fidelity is often measured by the training and competence of the interventionists and the degree of adherence to the intervention protocol. Ongoing supervision of interventionists and remediation of the intervention if it is not being delivered competently are critical to ensuring that minimum levels of fidelity are maintained (Carroll and Rounsaville, 2007; Miller et al., 2005). Established measures of adherence and competence (e.g., the Yale Adherence and Competence Scale; Carroll et al., 2000) can be adapted for use with newer interventions (Chawla et al., 2010). In pharmacotherapy trials, fidelity is often indexed by adherence to medication, measured using self-report, pill counts, urinary riboflavin, electronic medication event monitoring systems (MEMS), or monitoring of medication blood levels. The degree of medication adherence can have an impact on alcohol treatment outcomes (Baros et al., 2007; Stout et al., 2014), as can the quality of the adherence monitoring (Swift et al., 2011). Thus, adherence rates and monitoring method should always be reported (Witkiewitz et al., this issue). Each of the adherence monitoring

techniques has strengths and weaknesses. For example, monitoring medication blood levels cannot be used to monitor adherence in the placebo arm of a study. Using observation of medication ingestion to monitor adherence is not always feasible and can be thwarted by subjects hiding the medication in their cheeks or under their tongues.

Timing of Assessments and Mechanisms of Behavior Change

The timing of assessments

In many alcohol clinical trials, assessments coincide with treatment visits, with one or multiple post-treatment assessments occurring over a follow-up period. Thus, the intervention duration, which ranges from a few weeks to several months, often determines the assessment schedule during treatment. Follow-up assessments, when conducted, are often done at two- or three-month intervals, generally for six to 12 months post-treatment. The choice of when to assess progress in treatment should take into account the expected process of change during treatment and the clinical course of AUD (Collins and Graham, 2002; Maisto et al., 2014). For example, in a recent alcohol clinical trial the greatest reduction in alcohol consumption occurred in the weeks preceding the initiation of treatment (Stasiewicz et al., 2013). Thus, including an assessment of pre-treatment drinking that is consistent with the primary drinking outcome (e.g., the percentage of heavy drinking days) and over a sufficient period prior to treatment to be a stable estimate may be critical to understanding the process of change in an alcohol clinical trial. Researchers who have limited resources may also consider a planned missingness approach, whereby individuals are randomly assigned to complete a limited number of assessment timepoints (Enders, 2010; Mun et al., 2015; Schafer and Graham, 2002). This approach can reduce costs, as well as participant and research staff burden, while also providing a wealth of data that can be analyzed using a variety of missing data approaches (described below).

With the growing availability of smartphones, the potential for repeated contacts via ecological momentary assessment (EMA) could be more widely incorporated into alcohol clinical trial designs, both for assessment and treatment support (Cohn et al., 2011; Gurvich et al.; Gustafson et al., 2014). EMA measures participants' current self-reported behaviors, environmental exposures, and experiences in real time, in the subjects' natural environments (Shiffman et al., 2008). The conduct of multiple assessments over time, either daily or multiple times within days, can be particularly useful for analyses involving time-ordered effects [e.g., mediation analyses (Kranzler et al., 2014; Miranda et al., 2014)] and for examining dynamic processes of change (Witkiewitz and Marlatt, 2004). The decision whether to use EMA should balance the impact of participant burden, which can be substantial, with the specific research questions that can be addressed uniquely using EMA. For example, if the primary outcome measure is any drinking during a specific period, then EMA may not be necessary (Shiffman et al., 2008).

Mechanisms and moderators of behavior change

One of the major advantages of multiple assessments is that they may allow an examination of the mechanisms of change that underlie treatment efficacy (Kazdin, 2007; Longabaugh and Magill, 2011; Longabaugh, 2007). However, examining mechanisms of change often

requires larger samples and raises unique measurement considerations. For example, it is important to measure the hypothesized mechanism on multiple occasions prior to the anticipated change in behavior and continue measurement until after the behavior is anticipated to change. Behavioral and *in vivo* assessments, such as EMA and laboratory-based studies, may be particularly useful to study hypothesized mechanisms of behavior change in alcohol treatment trials (Miranda et al., 2014).

The examination of potential mechanisms of change also requires specific analytic approaches. Such studies often require multiple longitudinal assessments, formal tests of mediation (Cheong et al., 2003; MacKinnon, 2008) and approaches that yield stronger causal inferences regarding mediators' effects on outcomes (MacKinnon and Pirlott, 2015).

Of growing interest in the context of personalized medicine, is the effort to identify genetic variants that moderate the response to treatment (e.g., Johnson et al., 2011; Kranzler et al., 2014). This approach promises to identify *a priori* individuals who are most likely to respond to a particular medication. From a clinical trials perspective, it also offers the potential for enrichment trials, in which participants are selected based upon their genotype, for which there is evidence of moderation. For a detailed discussion of this issue see Jones et al. (in press). In addition to identifying potential treatment responders, validated genetic moderators can help to elucidate the mechanism of a medication's effects by implicating a specific neurotransmitter receptor or other medication target or pathway.

Statistical Analyses

Data transformation and alternative distributions

Often, data from alcohol and drug treatment studies are not normally distributed. For example, participants in alcohol treatment trials generally report an increasing number of abstinent days during the active intervention, yielding highly skewed data with a preponderance of zero values (also called "zero inflation") or a preponderance of values at the highest end of the range of values (e.g., 100% days abstinent). Thus, it may be necessary to transform the data prior to analysis or use analytic approaches that accommodate non-normal data distributions. The most commonly applied data transformations in alcohol clinical trials, such as square-root or arcsine transformations (Project MATCH Research Group, 1997), do not eliminate the problem of zero-inflation because the square root and arcsine of zero are zero. Alternative statistical models, including generalized linear models (Atkins et al., 2013; DeSantis et al., 2013), growth mixture models (Gueorguieva et al., 2010; Witkiewitz and Masyn, 2008), and latent Markov models (Witkiewitz, 2008; Witkiewitz et al., 2010) may be useful to address such distributional problems. These alternative statistical models are also an area of active ongoing methodological research (Bauer and Curran, 2003; Morgan et al., 2014; Sher et al., 2011; Steinley and Brusco, 2011).

Incorporating site, therapist, and group/cohort effects

Methods that incorporate treatment site, therapist, and group/cohort effects in analyses are important for all alcohol clinical trials. Ignoring systematic effects of these factors can lead to substantial bias in treatment effect estimates and standard errors, with a loss of statistical power (Chu et al., 2011; Feaster et al., 2011; Kraemer and Robinson, 2005). Although the

inclusion of site, intervention condition, and site-by-treatment interaction effects as “fixed effects” (i.e., the inclusion of site indicator variables as a covariate in the analysis) has been recommended (Kraemer and Robinson, 2005), large site differences in sample size or treatment effects can nonetheless lead to biased standard errors and make interpretation of the findings difficult (Chu et al., 2011). Likewise, therapist effects often explain a considerable amount of the variance in psychosocial treatment outcomes (Imel et al., 2008; Moyers and Miller, 2013). Oftentimes in alcohol clinical trials research the therapists’ effects on the outcomes are larger than the specific effects for each active intervention being delivered (Miller and Moyers, 2015). Under some circumstances, problems may occur when fixed effects models are used to examine site, therapist, and group effects, e.g., when the sample size is unbalanced across sites, therapists, or groups. With sample size imbalance, the fixed effect estimate is weighted by the sample sizes within sites, therapists, or groups, i.e., the fixed effects will be more strongly influenced by the largest samples within the respective category. Finally, this approach is problematic when there is insufficient power to detect a site-by-treatment interaction due to small samples within sites or a small number of sites.

Given these limitations and the results of numerous simulation studies, many authors have recommended that multisite, multi-therapist, and multi-group studies be analyzed using mixed effects models (i.e., hierarchical linear models, multilevel models: Chu et al., 2011; Kahan and Morris, 2013; Kahan, 2014; Moerbeek et al., 2003). Such models incorporate site and/or therapist and/or group information as a random effect, allowing for an estimation of the variability in outcomes that are due to therapist or group differences within or between sites. Mixed effects models provide more efficient and less biased estimates of the treatment effect, less biased standard errors, and greater statistical power than models that ignore site, therapist, or group effects or that treat these effects as fixed (Chu et al., 2011; Kahan and Morris, 2013; Kahan, 2014; Moerbeek et al., 2003). The analysis of data from groups that allow for rolling admission (i.e., open groups) requires even greater care, as the non-independence of individuals within groups becomes a moving target (Morgan-Lopez and Fals-Stewart, 2006, 2007).

Likewise, we recognize that “site” and “therapist” can be moving targets in some treatment settings, so that it may not be feasible to incorporate site and therapist as random effects. For example, individuals within a trial may transfer between sites or be recruited from multiple diverse programs within a site (e.g., intensive outpatient, aftercare), and some individuals may see multiple therapists or be enrolled in multiple types of treatment with different therapists. In these situations it may be preferable to measure characteristics of the treatment received (e.g., number of groups attended) and/or therapists (e.g., working alliance with all providers) that could be incorporated as fixed effects in predicting treatment outcomes.

Primary outcome analyses

The primary analysis should be pre-specified in the clinical trial registration and based on the original allocation of participants to intervention conditions, such that all individuals initially allocated to an intervention group are included in the analyses as members of that group. This analysis, often referred to as an “intention-to-treat” (ITT) analysis, has been

compromised in some alcohol clinical trials by the use of a “modified” ITT approach or the incorrect use of the ITT approach (Del Re et al., 2013). The modified ITT approach inappropriately eliminates certain individuals from analyses either because of missing data (see section below) or because a minimum level of treatment adherence is not achieved, resulting in potentially biased treatment estimates. When study analyses require some level of adherence (called “as treated”) or total adherence (called “per protocol”), estimates of intervention effects will be biased unless non-adherence with the research protocol is random (Ye et al., 2014). Excluding randomized patients from the primary analysis has been considered acceptable when errors are made in the implementation of eligibility criteria or when patients never received any of the interventions (Fergusson et al., 2002). In cases of differential non-compliance across intervention groups, alternative approaches to ITT analyses, such as complier average causal effects models (Tucker et al., 2012; Ye et al., 2014), may produce less biased estimates of the intervention effects.

When a stratified randomization procedure is used, the stratification variables need to be reflected in the analysis and the resulting study report (Witkiewitz et al., this issue). Failure to account for balancing variables can lead to non-independence of subjects across treatment groups, reduced statistical power, biased standard errors, and increased type-I error rates (Kahan and Morris, 2012).

Subgroup (i.e., moderator) analyses

Rigorous (vs. exploratory) analyses of alcohol clinical trials that examine potential subgroup differences (van den Brink et al., 2013) or personalized medicine approaches (Kranzler and McKay, 2012) require pre-specification of the analyses and adequate power to support them. Measurement of treatment moderators may require additional assessments, such as in pharmacogenetic studies, which require the collection of blood, saliva, or other tissue samples from which DNA can be extracted for genotyping (Goldman et al., 2005; Johnson et al., 2011). Important elements in a rigorous subgroup analysis include the use of a valid randomization procedure, adjustment for covariates with correction for multiple comparisons, consideration of the potential for non-replicability of subgroup differences when overall intervention effect sizes are small or absent (Grouin et al., 2005), and evaluation of the statistical significance of subgroup differences via tests of moderation effects.

Moderation analyses are critical for advancing therapeutic development and personalized medicine (Kranzler and McKay, 2012), but are particularly susceptible to bias, especially when subgroups are small (Brookes et al., 2004). Investigators are encouraged to consider methods to estimate effect sizes of treatment moderators (Kraemer, 2013) and methods to combine moderation analyses across studies to increase sample size and reduce bias (Brown et al., 2013; Wallace et al., 2013).

In general, any tests of moderation should ideally be planned *a priori* based on prior research or theoretical considerations and reflected in the trial registration. Analyses that are proposed after examining the collected data may be important for scientific reasons, but may yield spurious findings (Freemantle, 2001; Pocock et al., 1987). When reporting such

findings, they should be identified as having resulted from analyses conducted after the data were collected (Witkiewitz et al., this issue).

Missing data analyses

Unfortunately, missing data are common in alcohol clinical trials. Even the most carefully designed and executed studies are likely to experience some participant dropout, which occurs for a variety of reasons (Ball et al., 2006; Brorson et al., 2013; Coulson et al., 2009; Palmer et al., 2009). Numerous causes of missing data have been identified, such as relocation, incarceration, changes in employment and childcare needs, though oftentimes it is not possible to determine why some data are missing. The first critical step in dealing with missing data is to conduct attrition analyses to examine whether participants with missing data and those with complete data are systematically different on any available measures that were obtained prior to dropout. Importantly, many studies might not be powered to show effects on attrition, even when systematic influences on attrition do exist.

Variables most critical to examine in such analyses include demographic and other baseline variables that were identified *a priori* as potentially impacting outcomes, and any outcome measures that were collected prior to the time of dropout from the study. Systematic attrition biases can also be assessed via sensitivity analyses that examine the impact of missing data on the outcomes of interest (Enders, 2011, 2010; Jackson et al., 2014; Schafer and Graham, 2002). Pretreatment variables that predict attrition should be accounted for in subsequent analyses (National Research Council, 2010). Investigators might also consider including auxiliary measures at baseline that have been shown to predict attrition in prior studies (e.g., conscientiousness, honesty, and humility; Satherley et al., 2015) or may also consider adding a question at each assessment regarding the participant's intention to stay enrolled in the study (Schafer and Graham, 2002). The inclusion of auxiliary variables (defined as measures that may be associated with attrition and not necessarily associated with the outcomes of interest) can be critical in the missing data analytic models described next (Graham, 2009).

The second step in dealing with missing data is to use an analytic model that produces the least biased estimates of the intervention effect, even in the presence of missing data. Witkiewitz and colleagues recently published two studies that compared commonly used statistical approaches for handling missing data for continuous outcomes in alcohol clinical trials, using a simulation approach (Hallgren and Witkiewitz, 2014) and real world data (Witkiewitz et al., 2014). The approaches included last observation carried forward, baseline observation carried forward, placebo mean imputation, poor outcomes (e.g., heavy drinking) imputation, full information maximum likelihood, and multiple imputation. Consistent with 30 years of prior research on missing data approaches, the findings from both studies clearly showed full information maximum likelihood and multiple imputation to generate the least-biased estimates of intervention effects in alcohol clinical trials. Importantly, in both studies, two of the most commonly used methods in alcohol clinical trials research—last observation carried forward and poor outcome imputation (e.g., assuming drinking; Del Re et al., 2013)—tended to produce the most biased estimates of treatment effects in alcohol clinical trials (Hallgren and Witkiewitz, 2014; Witkiewitz et al., 2014). Although they are the preferred

approaches, full information maximum likelihood estimation and multiple imputation can introduce bias if data are missing not at random. Sensitivity analyses to examine the effects of different missing data models can be useful in evaluating the impact of missing data on the analysis of primary outcomes (Enders, 2010, 2011; Jackson et al., 2014; Witkiewitz et al., 2012).

Conclusions and Future Directions

Calls to develop additional and more efficacious treatments for AUDs can be found in the peer-reviewed scientific literature (Anton et al., 2012; Davies et al., 2013; Litten et al., 2012; Witkiewitz and Marlatt, 2011), as well as the popular press (Beck, 2014). The need for more treatment options with greater efficacy is reflected in the substantial global public health impact and societal costs of heavy drinking (Bouchery et al., 2011; World Health Organization, 2011) and by the high proportion of individuals with AUDs who never receive treatment or who drop out from treatment before receiving benefits from it (Substance Abuse and Mental Health Services Administration, 2012).

A key factor in the development of more efficacious AUD treatments is the use of sound methods to evaluate treatment efficacy. Recent reviews have shown that many RCTs in alcohol treatment have produced potentially biased estimates of treatment effects due to faulty design and analysis decisions (Del Re et al., 2013; Hallgren and Witkiewitz, 2014; Humphreys et al., 2008; Jonas et al., 2014a). Some of these problems identified include excessive exclusion criteria, biased randomization, inappropriate primary outcome analyses, and the use of faulty approaches to handle missing data.

The current review focused on design and analysis considerations for RCTs and the evaluation of treatment efficacy, yet many of the recommendations are also suitable for comparative effectiveness RCTs. Comparative effectiveness studies and pragmatic trials are critical for informing clinical care, but a review of methods for comparative effectiveness research is beyond the scope of the current review. A recent review of the comparative effectiveness of pharmacotherapies for alcohol use disorder provides some suggestions for future research (Jonas et al., 2014b). Guidelines for pragmatic trials have also been developed (Loudon et al., 2015).

Based on our systematic review of the literature, we provide ten recommendations for future research in Table 2. Valid methods developed for use in treatment trials for a variety of disorders, including AUDs, are available for the design and implementation of alcohol treatment trials. The high cost of clinical trials (both in the resources expended and the exposure of participants to the risks of experimental treatments) argue forcefully for the use of state-of-the-art methods in treatment trials for the disorder. The availability of new, highly efficacious treatments that could result from the use of these methods could help to remedy the low rate of treatment utilization by individuals with an AUD.

Acknowledgments

This research was supported by grants from the National Institute on Alcohol Abuse and Alcoholism (AA022328 to K. Witkiewitz and AA021164 and AA023192 to H. Kranzler) and by the U.S. Department of Veteran Affairs

(RCS-14-232 and the VISN 4 Mental Illness Research, Education, and Clinical Center). The views expressed do not reflect those of the Department of Veteran Affairs or other institutions.

References

- Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche PC, Lang T. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001; 134:663–694. [PubMed: 11304107]
- Anton RE, Randall CL. Measurement and choice of drinking outcome variables in the COMBINE Study. *J Stud Alcohol Suppl*. 2005:104–9. discussion 92–3. [PubMed: 16223062]
- Anton RF. Obsessive-compulsive aspects of craving: development of the Obsessive Compulsive Drinking Scale. *Addiction*. 2000; 95(Suppl 2):S211–7. [PubMed: 11002915]
- Anton RF, Litten RZ, Falk DE, Palumbo JM, Bartus RT, Robinson RL, Kranzler HR, Kosten TR, Meyer RE, O'Brien CP, Mann K, Meulien D. The Alcohol Clinical Trials Initiative (ACTIVE): purpose and goals for assessing important and salient issues for medications development in alcohol use disorders. *Neuropsychopharmacology*. 2012; 37:402–11. [PubMed: 21900883]
- Anton RF, O'Malley SS, Ciraulo DA, Cisler RA, Couper D, Donovan DM, Gastfriend DR, Hosking JD, Johnson BA, LoCastro JS, Longabaugh R, Mason BJ, Mattson ME, Miller WR, Pettinati HM, Randall CL, Swift R, Weiss RD, Williams LD, Zweben A. Combined pharmacotherapies and behavioral interventions for alcohol dependence: the COMBINE study: a randomized controlled trial. *JAMA*. 2006; 295:2003–17. [PubMed: 16670409]
- Atkins DC, Baldwin SA, Zheng C, Gallop RJ, Neighbors C. A tutorial on count regression and zero-altered count models for longitudinal substance use data. *Psychol Addict Behav*. 2013; 27:166–77. [PubMed: 22905895]
- Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med*. 2008; 27:2037–49. [PubMed: 18038446]
- Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. *Stat Med*. 2013; 32:2837–49. [PubMed: 23239115]
- Bakhshi E, McArdle B, Mohammad K, Seifi B, Biglarian A. Let continuous outcome variables remain continuous. *Comput Math Methods Med*. 2012; 2012:639124. [PubMed: 22693539]
- Ball SA, Carroll KM, Canning-Ball M, Rounsaville BJ. Reasons for dropout from drug abuse treatment: symptoms, personality, and motivation. *Addict Behav*. 2006; 31:320–30. [PubMed: 15964152]
- Barnett NP, Meade EB, Glynn TR. Predictors of detection of alcohol use episodes using a transdermal alcohol sensor. *Exp Clin Psychopharmacol*. 2014; 22:86–96. [PubMed: 24490713]
- Barnett NP, Tidey J, Murphy JG, Swift R, Colby SM. Contingency management for alcohol use reduction: a pilot study using a transdermal alcohol sensor. *Drug Alcohol Depend*. 2011; 118:391–9. [PubMed: 21665385]
- Baros AM, Latham PK, Moak DH, Voronin K, Anton RF. What role does measuring medication compliance play in evaluating the efficacy of naltrexone? *Alcohol Clin Exp Res*. 2007; 31:596–603. [PubMed: 17374038]
- Bauer DJ, Curran PJ. Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychol Methods*. 2003; 8:338–63. [PubMed: 14596495]
- Beck M. A prescription to end drinking. *Wall Str J*. 2014
- Berger VW. Quantifying the magnitude of baseline covariate imbalances resulting from selection bias in randomized clinical trials. *Biometrical J*. 2005a; 47:119–27. discussion 128–39.
- Berger VW. The reverse propensity score to detect selection bias and correct for baseline imbalances. *Stat Med*. 2005b; 24:2777–87. [PubMed: 15981305]
- Berger VW. A review of methods for ensuring the comparability of comparison groups in randomized clinical trials. *Rev Recent Clin Trials*. 2006; 1:81–6. [PubMed: 18393784]
- Blodgett JC, Del Re AC, Maisel NC, Finney JW. A meta-analysis of topiramate's effects for individuals with alcohol use disorders. *Alcohol Clin Exp Res*. 2014; 38:1481–8. [PubMed: 24796492]

- Bouchery EE, Harwood HJ, Sacks JJ, Simon CJ, Brewer RD. Economic costs of excessive alcohol consumption in the U.S., 2006. *Am J Prev Med.* 2011; 41:516–24. [PubMed: 22011424]
- Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clinical Epidemiol.* 2004; 57:229–36.
- Brorson HH, Ajo Arnevik E, Rand-Hendriksen K, Duckert F. Drop-out from addiction treatment: a systematic review of risk factors. *Clin Psychol Rev.* 2013; 33:1010–24. [PubMed: 24029221]
- Brown CH, Sloboda Z, Faggiano F, Teasdale B, Keller F, Burkhart G, Vigna-Taglianti F, Howe G, Masyn K, Wang W, Muthén B, Stephens P, Grey S, Perrino T. Methods for synthesizing findings on moderation effects across multiple randomized trials. *Prev Sci.* 2013; 14:144–56. [PubMed: 21360061]
- Cacciola JS, Alterman AI, Habing B, McLellan AT. Recent status scores for version 6 of the Addiction Severity Index (ASI-6). *Addiction.* 2011; 106:1588–602. [PubMed: 21545666]
- 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; 87:297–302. [PubMed: 17045423]
- Carroll KM, Nich C, Sifry RL, Nuro KF, Frankforter TL, Ball SA, Fenton L, Rounsaville BJ. A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug Alcohol Depend.* 2000; 57:225–38. [PubMed: 10661673]
- Carroll KM, Rounsaville BJ. A vision of the next generation of behavioral therapies research in the addictions. *Addiction.* 2007; 102:850–62. discussion 863–9. [PubMed: 17523974]
- Chakraborty B, Murphy SA. Dynamic Treatment Regimes. *Annu Rev Stat Its Appl.* 2014; 1:447–464.
- Chawla N, Collin S, Bowen S, Hsu S, Grow J, Douglass A, Marlatt GA. The mindfulness-based relapse prevention adherence and competence scale: development, interrater reliability, and validity. *Psychother Res.* 2010; 20:388–97. [PubMed: 20204916]
- Cheong J, Mackinnon DP, Khoo ST. Investigation of Mediational Processes Using Parallel Process Latent Growth Curve Modeling. *Struct Equ Model.* 2003; 10:238.
- Chu R, Thabane L, Ma J, Holbrook A, Pullenayegum E, Devereaux PJ. Comparing methods to estimate treatment effects on a continuous outcome in multicentre randomized controlled trials: a simulation study. *BMC Med Res Methodol.* 2011; 11:21. [PubMed: 21338524]
- Cisler RA, Kivlahan DR, Donovan D, Mattson ME. Assessing nondrinking outcomes in combined pharmacotherapy and psychotherapy clinical trials for the treatment of alcohol dependence. *J Stud Alcohol Suppl.* 2005; 110–8. discussion 92–3. [PubMed: 16223063]
- Cisler RA, Zweben A. Development of a composite measure for assessing alcohol treatment outcome: operationalization and validation. *Alcohol Clin Exp Res.* 1999; 23:263–71. [PubMed: 10069555]
- Clifford PR, Davis CM. Alcohol treatment research assessment exposure: a critical review of the literature. *Psychol Addict Behav.* 2012; 26:773–81. [PubMed: 22905899]
- Clifford PR, Maisto SA, Davis CM. Alcohol treatment research assessment exposure subject reactivity effects: part I. Alcohol use and related consequences. *J Stud Alcohol Drugs.* 2007; 68:519–28. [PubMed: 17568955]
- Cohn AM, Hunter-Reel D, Hagman BT, Mitchell J. Promoting behavior change from alcohol use through mobile technology: the future of ecological momentary assessment. *Alcohol Clin Exp Res.* 2011; 35:2209–15. [PubMed: 21689119]
- Collins LM, Graham JW. The effect of the timing and spacing of observations in longitudinal studies of tobacco and other drug use: temporal design considerations. *Drug Alcohol Depend.* 2002; 68(Suppl 1):S85–96. [PubMed: 12324177]
- Connors GJ, Maisto SA. Drinking reports from collateral individuals. *Addiction.* 2003; 98(Suppl 2): 21–9. [PubMed: 14984239]
- Coulson C, Ng F, Geertsema M, Dodd S, Berk M. Client-reported reasons for non-engagement in drug and alcohol treatment. *Drug Alcohol Rev.* 2009; 28:372–8. [PubMed: 19594791]
- Crunelle CL, Yegles M, van Nuijs ALN, Covaci A, De Doncker M, Maudens KE, Sabbe B, Dom G, Lambert WE, Michielsen P, Neels H. Hair ethyl glucuronide levels as a marker for alcohol use and abuse: a review of the current state of the art. *Drug Alcohol Depend.* 2014; 134:1–11. [PubMed: 24239414]

- Davies DL, Bortolato M, Finn DA, Ramaker MJ, Barak S, Ron D, Liang J, Olsen RW. Recent advances in the discovery and preclinical testing of novel compounds for the prevention and/or treatment of alcohol use disorders. *Alcohol Clin Exp Res*. 2013; 37:8–15. [PubMed: 22671690]
- Davis LL, Broome ME, Cox RP. Maximizing retention in community-based clinical trials. *J Nurs Scholarsh*. 2002; 34:47–53. [PubMed: 11901967]
- Del Boca FK, Darkes J. The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction*. 2003; 98(Suppl 2):1–12. [PubMed: 14984237]
- Del Re AC, Maisel NC, Blodgett JC, Finney JW. Intention-to-treat analyses and missing data approaches in pharmacotherapy trials for alcohol use disorders. *BMJ Open*. 2013; 3:e003464.
- Derogatis LR. The Brief Symptom Inventory: An introductory report. *Psychol Med A J Res Psychiatry Allied Sci*. 1983; 13:595–605.
- DeSantis SM, Bandyopadhyay D, Baker NL, Randall PK, Anton RF, Prisciandaro JJ. Modeling longitudinal drinking data in clinical trials: an application to the COMBINE study. *Drug Alcohol Depend*. 2013; 132:244–50. [PubMed: 23566774]
- Donohue B, Azrin NH, Strada MJ, Silver NC, Teichner G, Murphy H. Psychometric evaluation of self- and collateral timeline follow-back reports of drug and alcohol use in a sample of drug-abusing and conduct-disordered adolescents and their parents. *Psychol Addict Behav*. 2004; 18:184–9. [PubMed: 15238061]
- Donovan DM. More mice or a better mouse trap? Reflections on primary outcome indices in illicit drug dependence treatment research. *Addiction*. 2012; 107:723–4. [PubMed: 22372701]
- Donovan DM, Dunn CW, Rivara FP, Jurkovich GJ, Ries RR, Gentilello LM. Comparison of trauma center patient self-reports and proxy reports on the Alcohol Use Identification Test (AUDIT). *J Trauma*. 2004; 56:873–82. [PubMed: 15187756]
- Donovan DM, Kivlahan DR, Doyle SR, Longabaugh R, Greenfield SF. Concurrent validity of the Alcohol Use Disorders Identification Test (AUDIT) and AUDIT zones in defining levels of severity among out-patients with alcohol dependence in the COMBINE study. *Addiction*. 2006; 101:1696–704. [PubMed: 17156168]
- Dougherty DM, Hill-Kapturczak N, Liang Y, Karns TE, Cates SE, Lake SL, Mullen J, Roache JD. Use of continuous transdermal alcohol monitoring during a contingency management procedure to reduce excessive alcohol use. *Drug Alcohol Depend*. 2014; 142:301–6. [PubMed: 25064019]
- Enders, CK. *Applied missing data analysis*. Guilford Press; New York, NY: 2010.
- Enders CK. Missing not at random models for latent growth curve analyses. *Psychol Methods*. 2011; 16:1–16. [PubMed: 21381816]
- 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; 34:2022–34. [PubMed: 20659066]
- Fazzino TL, Harder VS, Rose GL, Helzer JE. A Daily Process Examination of the Bidirectional Relationship Between Craving and Alcohol Consumption as Measured Via Interactive Voice Response. *Alcohol Clin Exp Res*. 2013
- Feaster DJ, Mikulich-Gilbertson S, Brincks AM. Modeling site effects in the design and analysis of multi-site trials. *Am J Drug Alcohol Abuse*. 2011; 37:383–91. [PubMed: 21854281]
- Feinn R, Tennen H, Kranzler HR. Psychometric properties of the short index of problems as a measure of recent alcohol-related problems. *Alcohol Clin Exp Res*. 2003; 27:1436–41. [PubMed: 14506404]
- Fergusson D, Aaron SD, Guyatt G, Hébert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ*. 2002; 325:652–4. [PubMed: 12242181]
- Finney JW, Moyer A, Swearingen CE. Outcome variables and their assessment in alcohol treatment studies: 1968–1998. *Alcohol Clin Exp Res*. 2003; 27:1671–9. [PubMed: 14574240]
- Food and Drug Administration. *Statistical principles for clinical trials*. Center for Drug Evaluation and Research, Food and Drug Administration; Rockville, MD: 1998. Guidance for industry E9.
- Forcehimes AA, Tonigan JS, Miller WR, Kenna GA, Baer JS. Psychometrics of the Drinker Inventory of Consequences (DrInC). *Addict Behav*. 2007; 32:1699–704. [PubMed: 17182194]
- Freemantle N. Interpreting the results of secondary end points and subgroup analyses in clinical trials: should we lock the crazy aunt in the attic? *BMJ*. 2001; 322:989–91. [PubMed: 11312237]

- Gastfriend DR, Donovan D, Lefebvre R, Murray KT. Developing a baseline assessment battery: balancing patient time burden with essential clinical and research monitoring. *J Stud Alcohol Suppl.* 2005;94–103. discussion 92–3. [PubMed: 16223061]
- Goldman D, Oroszi G, O'Malley S, Anton R. COMBINE genetics study: the pharmacogenetics of alcoholism treatment response: genes and mechanisms. *J Stud Alcohol Suppl.* 2005;56–64. discussion 33. [PubMed: 16223057]
- Graham JW. Missing data analysis: making it work in the real world. *Annu Rev Psychol.* 2009; 60:549–76. [PubMed: 18652544]
- Grouin JM, Coste M, Lewis J. Subgroup analyses in randomized clinical trials: statistical and regulatory issues. *J Biopharm Stat.* 2005; 15:869–82. [PubMed: 16078390]
- Georgiueva R, Wu R, Donovan D, Rounsaville BJ, Couper D, Krystal JH, O'Malley SS. Naltrexone and combined behavioral intervention effects on trajectories of drinking in the COMBINE study. *Drug Alcohol Depend.* 2010; 107:221–9. [PubMed: 19969427]
- Georgiueva R, Wu R, Donovan D, Rounsaville BJ, Couper D, Krystal JH, O'Malley SS. Baseline trajectories of heavy drinking and their effects on postrandomization drinking in the COMBINE Study: empirically derived predictors of drinking outcomes during treatment. *Alcohol.* 2012; 46:121–31. [PubMed: 21925828]
- Gurvich EM, Kenna GA, Leggio L. Use of novel technology-based techniques to improve alcohol-related outcomes in clinical trials. *Alcohol Alcohol.* 48:712–9. [PubMed: 23955872]
- Gustafson DH, McTavish FM, Chih MY, Atwood AK, Johnson RA, Boyle MG, Levy MS, Driscoll H, Chisholm SM, Dillenburg L, Isham A, Shah D. A smartphone application to support recovery from alcoholism: a randomized clinical trial. *JAMA Psychiatry.* 2014; 71:566–72. [PubMed: 24671165]
- Hallgren, KA.; Atkins, D.; Witkiewitz, K. Whether to aggregate and how to analyze data in alcohol clinical trials: A comparison of type-I errors, power, and bias. under review
- Hallgren KA, Witkiewitz K. Missing data in alcohol clinical trials: A comparison of methods. *Alcohol Clin Exp Res.* 2014; 38
- Harris AHS, Bowe T, Finney JW, Humphreys K. HEDIS initiation and engagement quality measures of substance use disorder care: impact of setting and health care specialty. *Popul Health Manag.* 2009a; 12:191–6. [PubMed: 19663621]
- Harris AHS, Oliva E, Bowe T, Humphreys KN, Kivlahan DR, Trafton JA. Pharmacotherapy of alcohol use disorders by the Veterans Health Administration: patterns of receipt and persistence. *Psychiatr Serv.* 2012; 63:679–85. [PubMed: 22549276]
- Harris AHS, Reeder R, Hyun JK. Common statistical and research design problems in manuscripts submitted to high-impact psychiatry journals: what editors and reviewers want authors to know. *J Psychiatr Res.* 2009b; 43:1231–4. [PubMed: 19435635]
- Hartmann S, Aradottir S, Graf M, Wiesbeck G, Lesch O, Ramskogler K, Wolfersdorf M, Alling C, Wurst FM. Phosphatidylethanol as a sensitive and specific biomarker: comparison with gamma-glutamyl transpeptidase, mean corpuscular volume and carbohydrate-deficient transferrin. *Addict Bol.* 2007; 12:81–4.
- Hashimoto E, Riederer PF, Hesselbrock VM, Hesselbrock MN, Mann K, Ukai W, Sohma H, Thibaut F, Schuckit MA, Saito T. Consensus paper of the WFSBP task force on biological markers: biological markers for alcoholism. *World J Biol Psychiatry.* 2013; 14:549–64. [PubMed: 24236956]
- Heather N, Tebbutt JS, Mattick RP, Zamir R. Development of a scale for measuring impaired control over alcohol consumption: a preliminary report. *J Stud Alcohol.* 1993; 54:700–9. [PubMed: 8271806]
- Hedden SL, Woolson RF, Malcolm RJ. Randomization in substance abuse clinical trials. *Subst Abuse Treat Prev Policy.* 2006; 1:6. [PubMed: 16722591]
- Hinshaw SP, Hoagwood K, Jensen PS, Kratochvil C, Bickman L, Clarke G, Abikoff HB, Atkins M, Vitiello B. AACAP 2001 research forum: challenges and recommendations regarding recruitment and retention of participants in research investigations. *J Am Acad Child Adolesc Psychiatry.* 2004; 43:1037–45. [PubMed: 15266200]

- Hock B, Schwarz M, Domke I, Grunert VP, Wuertemberger M, Schiemann U, Horster S, Limmer C, Stecker G, Soyka M. Validity of carbohydrate-deficient transferrin (%CDT), gamma-glutamyltransferase (gamma-GT) and mean corpuscular erythrocyte volume (MCV) as biomarkers for chronic alcohol abuse: a study in patients with alcohol dependence and liver disorders of non-alcohol. *Addiction*. 2005; 100:1477–86. [PubMed: 16185209]
- Hoertel N, Falissard B, Humphreys K, Gorwood P, Seigneurie AS, Limosin F. Do clinical trials of treatment of alcohol dependence adequately enroll participants with co-occurring independent mood and anxiety disorders? An analysis of data from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *J Clin Psychiatry*. 2014; 75:231–7. [PubMed: 24569017]
- Humphreys K, Blodgett JC, Wagner TH. Estimating the Efficacy of Alcoholics Anonymous without Self-Selection Bias: An Instrumental Variables Re-Analysis of Randomized Clinical Trials. *Alcohol Clin Exp Res*. 2014; 38:2688–2694. [PubMed: 25421504]
- Humphreys K, Harris AHS, Weingardt KR. Subject eligibility criteria can substantially influence the results of alcohol-treatment outcome research. *J Stud Alcohol Drugs*. 2008; 69:757–64. [PubMed: 18781251]
- Humphreys K, Weisner CM. Use of exclusion criteria in selecting research subjects and its effect on the generalizability of alcohol treatment outcome studies. *Am J Psychiatry*. 2000; 157:588–94. [PubMed: 10739418]
- Imel ZE, Wampold BE, Miller SD, Fleming RR. Distinctions without a difference: direct comparisons of psychotherapies for alcohol use disorders. *Psychol Addict Behav*. 2008; 22:533–43. [PubMed: 19071978]
- Jackson D, White IR, Mason D, Sutton S. A general method for handling missing binary outcome data in randomized controlled trials. *Addiction*. 2014; 109:1986–93. [PubMed: 25171441]
- Jatlow PI, Agro A, Wu R, Nadim H, Toll BA, Ralevski E, Nogueira C, Shi J, Dziura JD, Petrakis IL, O'Malley SS. Ethyl glucuronide and ethyl sulfate assays in clinical trials, interpretation, and limitations: results of a dose ranging alcohol challenge study and 2 clinical trials. *Alcohol Clin Exp Res*. 2014; 38:2056–65. [PubMed: 24773137]
- Johnson BA, Ait-Daoud N, Seneviratne C, Roache JD, Javors MA, Wang XQ, Liu L, Penberthy JK, DiClemente CC, Li MD. Pharmacogenetic approach at the serotonin transporter gene as a method of reducing the severity of alcohol drinking. *Am J Psychiatry*. 2011; 168:265–75. [PubMed: 21247998]
- Jonas DE, Amick HR, Feltner C, Bobashev G, Thomas K, Wines R, Kim MM, Shanahan E, Gass CE, Rowe CJ, Garbutt JC. Pharmacotherapy for adults with alcohol use disorders in outpatient settings: a systematic review and meta-analysis. *JAMA*. 2014a; 311:1889–900. [PubMed: 24825644]
- Jonas, DE.; Amick, HR.; Feltner, C.; Bobashev, G.; Thomas, K.; Wines, R.; Kim, MM.; Shanahan, E.; Gass, CE.; Rowe, CJ.; Garbutt, JC. Comparative effectiveness review no. 134. Agency for Healthcare Research and Quality; Rockville, MD: 2014b. Pharmacotherapy for adults with alcohol-use disorders in outpatient settings.
- Jones, JD.; Comer, SD.; Kranzler, HR. [accessed 3.2.15] The pharmacogenetics of alcohol use disorder [WWW Document]. *Alcohol Clin Exp Res*. n.d. <http://www.ncbi.nlm.nih.gov/pubmed/25703505>
- Kahan BC. Accounting for centre-effects in multicentre trials with a binary outcome - when, why, and how? *BMC Med Res Methodol*. 2014; 14:20. [PubMed: 24512175]
- Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. *Stat Med*. 2012; 31:328–40. [PubMed: 22139891]
- Kahan BC, Morris TP. Analysis of multicentre trials with continuous outcomes: when and how should we account for centre effects? *Stat Med*. 2013; 32:1136–49. [PubMed: 23112128]
- Kaminer Y, Burleson JA, Burke R. Can assessment reactivity predict treatment outcome among adolescents with alcohol and other substance use disorders? *Subst Abus*. 2008; 29:63–9. [PubMed: 19042325]
- Kazdin AE. Mediators and mechanisms of change in psychotherapy research. *Annu Rev Clin Psychol*. 2007; 3:1–27. [PubMed: 17716046]

- Kazdin AE. Evidence-based treatment and practice: new opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *Am Psychol.* 2008; 63:146–59. [PubMed: 18377105]
- Kiluk BD, Dreifuss JA, Weiss RD, Morgenstern J, Carroll KM. The Short Inventory of Problems - revised (SIP-R): psychometric properties within a large, diverse sample of substance use disorder treatment seekers. *Psychol Addict Behav.* 2013; 27:307–14. [PubMed: 22642856]
- Kraemer HC. Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: a parametric approach. *Stat Med.* 2013; 32:1964–73. [PubMed: 23303653]
- Kraemer HC, Robinson TN. Are certain multicenter randomized clinical trial structures misleading clinical and policy decisions? *Contemp Clin Trials.* 2005; 26:518–29. [PubMed: 15993653]
- Kranzler HR, Armeli S, Tennen H, Covault J, Feinn R, Arias AJ, Pettinati H, Oncken C. A double-blind, randomized trial of sertraline for alcohol dependence: moderation by age of onset [corrected] and 5-hydroxytryptamine transporter-linked promoter region genotype. *J Clin Psychopharmacol.* 2011; 31:22–30. [PubMed: 21192139]
- Kranzler HR, Armeli S, Wetherill R, Feinn R, Tennen H, Gelernter J, Covault J, Pond T. Self-efficacy mediates the effects of topiramate and GRIK1 genotype on drinking. *Addict Biol.*
- Kranzler HR, Escobar R, Lee DK, Meza E. Elevated rates of early discontinuation from pharmacotherapy trials in alcoholics and drug abusers. *Alcohol Clin Exp Res.* 1996; 20:16–20. [PubMed: 8651447]
- Kranzler HR, McKay JR. Personalized treatment of alcohol dependence. *Curr Psychiatry Rep.* 2012; 14:486–93. [PubMed: 22810115]
- Kurtz SP, Surratt HL, Buttram ME, Levi-Minzi MA, Chen M. Interview as intervention: the case of young adult multidrug users in the club scene. *J Subst Abuse Treat.* 2013; 44:301–8. [PubMed: 22971689]
- Laforge RG, Borsari B, Baer JS. The utility of collateral informant assessment in college alcohol research: results from a longitudinal prevention trial. *J Stud Alcohol.* 2005; 66:479–87. [PubMed: 16240555]
- Lane SP, Sher KJ. Limits of current approaches to diagnosis severity based on criterion counts: An example with DSM-5 alcohol use disorder. *Clin Psychol Sci.* n.d
- Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A “SMART” design for building individualized treatment sequences. *Annu Rev Clin Psychol.* 2012; 8:21–48. [PubMed: 22224838]
- Leyrat C, Caille A, Donner A, Giraudeau B. Propensity scores used for analysis of cluster randomized trials with selection bias: a simulation study. *Stat Med.* 2013; 32:3357–72. [PubMed: 23553813]
- Litten RZ, Bradley AM, Moss HB. Alcohol biomarkers in applied settings: recent advances and future research opportunities. *Alcohol Clin Exp Res.* 2010; 34:955–67. [PubMed: 20374219]
- Litten RZ, Castle IJP, Falk D, Ryan M, Fertig J, Chen CM, Yi H. The placebo effect in clinical trials for alcohol dependence: an exploratory analysis of 51 naltrexone and acamprosate studies. *Alcohol Clin Exp Res.* 2013; 37:2128–37. [PubMed: 23889231]
- Litten RZ, Egli M, Heilig M, Cui C, Fertig JB, Ryan ML, Falk DE, Moss H, Huebner R, Noronha A. Medications development to treat alcohol dependence: a vision for the next decade. *Addict Biol.* 2012; 17:513–27. [PubMed: 22458728]
- Litten RZ, Ryan ML, Falk DE, Reilly M, Fertig JB, Koob GF. Heterogeneity of Alcohol Use Disorder: Understanding Mechanisms to Advance Personalized Treatment. *Alcohol Clin Exp Res.* 2015; 39:579–584. [PubMed: 25833016]
- LoCastro JS, Youngblood M, Cisler RA, Mattson ME, Zweben A, Anton RF, Donovan DM. Alcohol treatment effects on secondary nondrinking outcomes and quality of life: the COMBINE study. *J Stud Alcohol Drugs.* 2009; 70:186–96. [PubMed: 19261230]
- Longabaugh R. The search for mechanisms of change in behavioral treatments for alcohol use disorders: a commentary. *Alcohol Clin Exp Res.* 2007; 31:21s–32s. [PubMed: 17880343]
- Longabaugh R, Magill M. Recent advances in behavioral addiction treatments: focusing on mechanisms of change. *Curr Psychiatry Rep.* 2011; 13:382–9. [PubMed: 21750958]

- Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ*. 2015; 350:h2147. [PubMed: 25956159]
- MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. *Psychol Methods*. 2002; 7:19–40. [PubMed: 11928888]
- MacKinnon, DP. *Introduction to Statistical Mediation Analysis*. Taylor & Francis Group; New York: 2008.
- MacKinnon DP, Pirlott AG. Statistical Approaches for Enhancing Causal Interpretation of the M to Y Relation in Mediation Analysis. *Personal Soc Psychol Rev*. 2014; 19:30–43.
- Magura S, McKean J, Kosten S, Tonigan JS. A novel application of propensity score matching to estimate Alcoholics Anonymous' effect on drinking outcomes. *Drug Acohol Depend*. 2013; 129:54–9.
- Maisto SA, Clifford PR, Davis CM. Alcohol treatment research assessment exposure subject reactivity effects: part II. Treatment engagement and involvement. *J Stud Alcohol Drugs*. 2007; 68:529–33. [PubMed: 17568956]
- Maisto SA, Conigliaro JC, Gordon AJ, McGinnis KA, Justice AC. An experimental study of the agreement of self-administration and telephone administration of the Timeline Followback interview. *J Stud Alcohol Drugs*. 2008; 69:468–71. [PubMed: 18432391]
- Maisto SA, Kirouac M, Witkiewitz K. Alcohol use disorder clinical course research: informing clinicians' treatment planning now and in the future. *J Stud Alcohol Drugs*. 2014; 75:799–807. [PubMed: 25208198]
- Mann K, Bladström A, Torup L, Gual A, van den Brink W. Extending the treatment options in alcohol dependence: a randomized controlled study of as-needed nalmefene. *Biol Psychiatry*. 2013; 73:706–13. [PubMed: 23237314]
- Matts JP, Lachin JM. Properties of permuted-block randomization in clinical trials. *Control Clin Trials*. 1988; 9:327–44. [PubMed: 3203524]
- McCaffrey DF, Griffin BA, Almirall D, Slaughter ME, Ramchand R, Burgette LF. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Stat Med*. 2013; 32:3388–414. [PubMed: 23508673]
- McKay, JR. *Treating substance use disorders with adaptive continuing care*. American Psychological Association; Washington, DC US: 2009. *Treating substance use disorders with adaptive continuing care*.
- McLellan, aT; Kushner, H.; Metzger, D.; Peters, R.; Smith, I.; Grissom, G.; Pettinati, H.; Argeriou, M. The Fifth Edition of the Addiction Severity Index. *J Subst Abuse Treat*. 1992; 9:199–213. [PubMed: 1334156]
- Merkx MJM, Schippers GM, Koeter MWJ, De Wildt WAJM, Vedel E, Goudriaan AE, Van Den Brink W. Treatment outcome of alcohol use disorder outpatients with or without medically assisted detoxification. *J Stud Alcohol Drugs*. 2014; 75:993–8. [PubMed: 25343657]
- Miller, WR. *Project MA. Form 90: A structured assessment interview for drinking and related behaviors*. National Institute on Alcohol Abuse and Alcoholism; Bethesda, MD: 1996.
- Miller WR, Moyers TB. The forest and the trees: relational and specific factors in addiction treatment. *Addiction*. 2014
- Miller WR, Moyers TB, Arciniega L, Ernst D, Forcehimes A. Training, supervision and quality monitoring of the COMBINE Study behavioral interventions. *J Stud Alcohol Suppl*. 2005:188–195. [PubMed: 16223070]
- Miller, WR.; Tonigan, JS.; Longabaugh, R. *Project MA. The Drinker Inventory of Consequences (DrInC)*. National Institute on Alcohol Abuse and Alcoholism; Bethesda, MD: 1995.
- Miller WR, Wilbourne PL. Mesa Grande: a methodological analysis of clinical trials of treatments for alcohol use disorders. *Addiction*. 2002; 97:265–277. [PubMed: 11964100]
- Miranda R, MacKillop J, Treloar H, Blanchard A, Tidey JW, Swift RM, Chun T, Rohsenow DJ, Monti PM. Biobehavioral mechanisms of topiramate's effects on alcohol use: an investigation pairing laboratory and ecological momentary assessments. *Addict Biol*.
- Moerbeek M, van Breukelen GJP, Berger MPF. A comparison between traditional methods and multilevel regression for the analysis of multicenter intervention studies. *J Clin Epidemiol*. 2003; 56:341–50. [PubMed: 12767411]

- Morgan CJ, Lenzenweger MF, Rubin DB, Levy DL. A hierarchical finite mixture model that accommodates zero-inflated counts, non-independence, and heterogeneity. *Stat Med*. 2014; 33:2238–50. [PubMed: 24443287]
- Morgan-Lopez AA, Fals-Stewart W. Analytic complexities associated with group therapy in substance abuse treatment research: problems, recommendations, and future directions. *Exp Clin Psychopharmacol*. 2006; 14:265–73. [PubMed: 16756430]
- Morgan-Lopez AA, Fals-Stewart W. Analytic methods for modeling longitudinal data from rolling therapy groups with membership turnover. *J Consult Clin Psychol*. 2007; 75:580–93. [PubMed: 17663612]
- Moyers TB, Miller WR. Is low therapist empathy toxic? *Psychol Addict Behav*. 2013; 27:878–84. [PubMed: 23025709]
- Mun EY, de la Torre J, Atkins DC, White HR, Ray AE, Kim SY, Jiao Y, Clarke N, Huo Y, Larimer ME, Huh D. Project INTEGRATE: An integrative study of brief alcohol interventions for college students. *Psychol Addict Behav*. 2015; 29:34–48. [PubMed: 25546144]
- Murphy SA, Lynch KG, Oslin D, McKay JR, TenHave T. Developing adaptive treatment strategies in substance abuse research. *Drug Alcohol Depend*. 2007; 88(Suppl 2):S24–30. [PubMed: 17056207]
- Myers JA, Rassen JA, Gagne JJ, Huybrechts KF, Schneeweiss S, Rothman KJ, Joffe MM, Glynn RJ. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am J Epidemiol*. 2011; 174:1213–22. [PubMed: 22025356]
- Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, Fabiano GA, Waxmonsky JG, Yu J, Murphy SA. Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychol Methods*. 2012; 17:457–77. [PubMed: 23025433]
- National Institute for Health and Care Excellence. Alcohol-use disorders: Diagnosis, assessment and management of harmful drinking and alcohol dependence. National Institute for Health and Care Excellence; Manchester, UK: 2011.
- National Research Council. The prevention and treatment of missing data in clinical trials. The National Academies Press; Washington, DC, US: 2010.
- Newnham EA, Page AC. Bridging the gap between best evidence and best practice in mental health. *Clin Psychol Rev*. 2010; 30:127–42. [PubMed: 19897295]
- Palmer RS, Murphy MK, Piselli A, Ball SA. Substance user treatment dropout from client and clinician perspectives: a pilot study. *Subst Use Misuse*. 2009; 44:1021–38. [PubMed: 19938942]
- Pedersen ER, Grow J, Duncan S, Neighbors C, Larimer ME. Concurrent validity of an online version of the Timeline Followback assessment. *Psychol Addict Behav*. 2012; 26:672–7. [PubMed: 22486334]
- Pedersen ER, LaBrie JW. A within-subjects validation of a group-administered timeline followback for alcohol use. *J Stud Alcohol*. 2006; 67:332–5. [PubMed: 16562417]
- Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med*. 1987; 317:426–32. [PubMed: 3614286]
- Project MATCH Research Group. Matching Alcoholism Treatments to Client Heterogeneity: Project MATCH posttreatment drinking outcomes. *J Stud Alcohol*. 1997; 58:7–29. [PubMed: 8979210]
- Robinson SM, Sobell LC, Sobell MB, Arcidiacono S, Tzall D. Alcohol and drug treatment outcome studies: new methodological review (2005–2010) and comparison with past reviews. *Addict Behav*. 2014; 39:39–47. [PubMed: 24140306]
- Ruta D, Garratt A, Abdalla M, Buckingham K, Russell I. The SF 36 health survey questionnaire. A valid measure of health status.... *BMJ*. 1993; 307:448–9. [PubMed: 8374477]
- Rychtarik RG, McGillicuddy NB, Connors GJ, Whitney RB. Participant selection biases in a randomized clinical trial of alcoholism treatment settings and intensities. *Alcohol Clin Exp Res*. 1998; 22:969–73. [PubMed: 9726264]
- Satherley N, Milojev P, Greaves LM, Huang Y, Osborne D, Bulbulia J, Sibley CG. Demographic and psychological predictors of panel attrition: evidence from the new zealand attitudes and values study. *PLoS One*. 2015; 10:e0121950. [PubMed: 25793746]
- Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods*. 2002; 7:147–77. [PubMed: 12090408]

- Sher KJ, Jackson KM, Steinley D. Alcohol use trajectories and the ubiquitous cat's cradle: cause for concern? *J Abnorm Psychol.* 2011; 120:322–35. [PubMed: 21319874]
- Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol.* 2008; 4:1–32. [PubMed: 18509902]
- Sobell LC, Agrawal S, Sobell MB. Factors affecting agreement between alcohol abusers' and their collaterals' reports. *J Stud Alcohol.* 1997; 58:405–13. [PubMed: 9203122]
- Sobell LC, Brown J, Leo GI, Sobell MB. The reliability of the Alcohol Timeline Followback when administered by telephone and by computer. *Drug Alcohol Depend.* 1996; 42:49–54. [PubMed: 8889403]
- Sobell, LC.; Sobell, MB. Timeline Follow-Back: A technique for assessing self-reported alcohol consumption. Human Press; Totowa, NJ, US: 1992.
- Stasiewicz PR, Schlauch RC, Bradizza CM, Bole CW, Coffey SF. Pretreatment changes in drinking: relationship to treatment outcomes. *Psychol Addict Behav.* 2013; 27:1159–66. [PubMed: 23276313]
- Steinley D, Brusco MJ. Evaluating mixture modeling for clustering: recommendations and cautions. *Psychol Methods.* 2011; 16:63–79. [PubMed: 21319900]
- Stout RL, Braciszewski JM, Subbaraman MS, Kranzler HR, O'Malley SS, Falk D. What happens when people discontinue taking medications? Lessons from COMBINE Addiction. 2014; 109:2044–52. [PubMed: 25098969]
- Substance Abuse and Mental Health Services Administration. Results from the 2010 National Survey on Drug Use and Health: Summary of National Findings. Rockville, MD: 2012.
- Substance Abuse and Mental Health Services Administration. NSDUH Seri. Results from the 2013 National Survey on Drug Use and Health: Summary of National Findings. Substance Abuse and Mental Health Services Administration; Rockville, MD: 2014.
- Swift R, Oslin DW, Alexander M, Forman R. Adherence monitoring in naltrexone pharmacotherapy trials: a systematic review. *J Stud Alcohol Drugs.* 2011; 72:1012–8. [PubMed: 22051215]
- 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.* 2012a; 107:709–18. [PubMed: 21981638]
- Tiffany ST, Friedman L, Greenfield SF, Hasin DS, Jackson R. Other outcomes in treatments for substance-use disorders: a call for action. *Addiction.* 2012b; 107:709–18. [PubMed: 21981638]
- Tucker JA, Roth DL, Huang J, Crawford MS, Simpson CA. Effects of interactive voice response self-monitoring on natural resolution of drinking problems: utilization and behavioral economic factors. *J Stud Alcohol Drugs.* 2012; 73:686–98. [PubMed: 22630807]
- U.S. Burden of Disease Collaborators. The state of US health, 1990–2010: burden of diseases, injuries, and risk factors. *JAMA.* 2013; 310:591–608. [PubMed: 23842577]
- UKATT Research Team. Effectiveness of treatment for alcohol problems: findings of the randomised UK alcohol treatment trial (UKATT). *BMJ.* 2005; 331:541. [PubMed: 16150764]
- Van den Brink W, Aubin HJ, Bladström A, Torup L, Gual A, Mann K. Efficacy of as-needed nalmefene in alcohol-dependent patients with at least a high drinking risk level: results from a subgroup analysis of two randomized controlled 6-month studies. *Alcohol Alcohol.* 2013; 48:570–8. [PubMed: 23873853]
- Wallace ML, Frank E, Kraemer HC. A novel approach for developing and interpreting treatment moderator profiles in randomized clinical trials. *JAMA Psychiatry.* 2013; 70:1241–7. [PubMed: 24048258]
- Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* 1996; 34:220–33. [PubMed: 8628042]
- Weiss RD, O'malley SS, Hosking JD, Locastro JS, Swift R. Do patients with alcohol dependence respond to placebo? Results from the COMBINE Study. *J Stud Alcohol Drugs.* 2008; 69:878–84. [PubMed: 18925346]
- Whitford JL, Widner SC, Mellick D, Elkins RL. Self-report of drinking compared to objective markers of alcohol consumption. *Am J Drug Alcohol Abuse.* 2009; 35:55–8. [PubMed: 19322728]
- Witkiewitz K. Lapses following alcohol treatment: modeling the falls from the wagon. *J Stud Alcohol Drugs.* 2008; 69:594–604. [PubMed: 18612576]

- Witkiewitz K. "Success" following alcohol treatment: moving beyond abstinence. *Alcohol Clin Exp Res*. 2013; 37(Suppl 1):E9–13. [PubMed: 23075307]
- Witkiewitz K, Bush T, Magnusson LB, Carlini BH, Zbikowski SM. Trajectories of cigarettes per day during the course of telephone tobacco cessation counseling services: a comparison of missing data models. *Nicotine Tob Res*. 2012; 14:1100–4. [PubMed: 22180594]
- Witkiewitz K, Falk DE, Kranzler HR, Litten RZ, Hallgren KA, O'Malley SS, Anton RF. Methods to analyze treatment effects in the presence of missing data for a continuous heavy drinking outcome measure when participants drop out from treatment in alcohol clinical trials. *Alcohol Clin Exp Res*. 2014; 38:2826–34. [PubMed: 25421518]
- Witkiewitz K, Finney J, Harris A, Kivlahan D, Kranzler H. Guidelines for the reporting of treatment trials for alcohol use disorders. *Alcohol Clin Exp Res*. This Issue.
- Witkiewitz K, Maisto SA, Donovan DM. A comparison of methods for estimating change in drinking following alcohol treatment. *Alcohol Clin Exp Res*. 2010; 34:2116–2125. [PubMed: 20860611]
- Witkiewitz K, Marlatt GA. Relapse Prevention for Alcohol and Drug Problems: That Was Zen, This Is Tao. *Am Psychol*. 2004; 59:224–235. [PubMed: 15149263]
- Witkiewitz K, Marlatt GA. Behavioral therapy across the spectrum. *Alcohol Res Heal*. 2011; 33:313–319.
- Witkiewitz K, Masyn KE. Drinking trajectories following an initial lapse. *Psychol Addict Behav*. 2008; 22:157–67. [PubMed: 18540713]
- World Health Organization. World Health Organization Quality of Life. World Health Organization; Geneva, Switzerland: 1998.
- World Health Organization. Global health risks: mortality and burden of disease attributable to selected major risks. World Health Organization; Geneva, Switzerland: 2009.
- World Health Organization. Global Status Report on Alcohol and Health. Geneva: 2011.
- Ye C, Beyene J, Browne G, Thabane L. Estimating treatment effects in randomised controlled trials with non-compliance: a simulation study. *BMJ Open*. 2014; 4:e005362.
- Ye Y, Kaskutas LA. Using propensity scores to adjust for selection bias when assessing the effectiveness of Alcoholics Anonymous in observational studies. *Drug Alcohol Depend*. 2009; 104:56–64.
- Yoo B. The impact of dichotomization in longitudinal data analysis: a simulation study. *Pharm Stat*. 2010; 9:298–312. [PubMed: 19904810]
- Zindel LR, Kranzler HR. Pharmacotherapy of alcohol use disorders: seventy-five years of progress. *J Stud Alcohol Drugs*. 2014; 75(Suppl 1):79–88.
- Zweben A, Barrett D, Berger L, Murray KT. Recruiting and retaining participants in a combined behavioral and pharmacological clinical trial. *J Stud Alcohol Suppl*. 2005a:72–81. discussion 65. [PubMed: 16223059]
- Zweben A, Barrett D, Berger L, Murray KT. Recruiting and retaining participants in a combined behavioral and pharmacological clinical trial. *J Stud Alcohol Suppl*. 2005b:72–81. discussion 65. [PubMed: 16223059]

Table 1

Summary of Methods for Randomization in RCTs.

Procedure	Description	Advantages	Disadvantages
Complete	Each successive patient has an equal probability of assignment to all treatments	Ease of use	Very high potential for imbalances; no covariate stratification
Permuted block	Within each successive block patients have a random probability of assignment to each treatment	Ease of use; stratification on covariates	Potential for imbalance; need to control for stratified covariates in analyses
Urn randomization	Adaptive randomization scheme where probability of assignment is adjusted based on degree of treatment of imbalance	Reduces imbalance; stratification on covariates	Difficulty in implementation; need to control for stratified covariates in analyses
Covariate adaptive	Adaptive randomization scheme where probability of assignment is adjusted based on degree of treatment of imbalance and to minimize covariate imbalance	Reduces treatment imbalance and covariate imbalance	Difficulty in implementation; need to control for covariates in analyses; assignment is not “random”
Two-stage approach	Combines urn randomization with covariate adaptive randomization	Reduces treatment imbalance and covariate imbalance	Not widely used; need to control for covariates in analyses; assignment is not “random”

Table 2

Summary of Recommendations for Design and Analysis of Alcohol Treatment Studies.

Recommendation	Selected Relevant Citations
1. Careful consideration and reporting of exclusion criteria	Altman et al., 2001; Humphreys et al., 2008
2. Selection of a randomization procedure that reduces imbalances across treatment groups	Hedden et al., 2006; Matts and Lachin, 1988
3. Consider use of propensity score or instrumental variable methods when group imbalances exist	Berger, 2005b; Leyrat et al., 2013
4. Use of well-validated self-report measures and either biomarkers or carefully selected collateral informants to validate self-reported consumption	Litten et al., 2010; Miller, 1996; Sobell and Sobell, 1992; Sobell et al., 1997
5. Careful consideration of the timing of assessments to inform the durability of treatment effects and variation in outcomes over time	Collins and Graham, 2002; Stasiewicz et al., 2013
6. Investigation of data distributions and the use of either data transformation or alternative analytic techniques when data are non-normal	Atkins et al., 2013; Gueorguieva et al., 2012; Witkiewitz et al., 2010
7. An intention-to-treat (ITT) analytic approach that takes into account clustering of patients within sites, therapists, and/or groups and that controls for covariate adjustment in the randomization procedure	Del Re et al., 2013; Kahan and Morris, 2013; Moerbeek et al., 2003
8. An alternative approach to ITT, such as the complier average causal effect model, if there is unequal compliance across groups	Tucker et al., 2012; Ye et al., 2014
9. Minimizing missing primary outcome data by continuing to assess all individuals (including those who drop out of treatment) and maximum likelihood estimation or multiple imputation to accommodate missing data	Enders, 2010; Hallgren and Witkiewitz, 2014; Witkiewitz et al., 2014
10. Sensitivity analyses to evaluate the impact of missing data on the treatment effect estimates	Enders, 2011; Jackson et al., 2014