Is Motivational Interviewing Effective at Reducing Alcohol Misuse in Young Adults?:

A Critical Review of Foxcroft et al. (2014)

Eun-Young Mun[1], **David C. Atkins**[2], and **Scott T. Walters**[3]

[1]Center of Alcohol Studies, Rutgers, The State University of New Jersey

[2]Department of Psychiatry and Behavioral Sciences, The University of Washington

[3]Department of Behavioral and Community Health, The University of North Texas Health Science Center

Abstract

Foxcroft, Coombes, Wood, Allen, and Almeida Santimano (2014) recently conducted a meta-analysis evaluating the effectiveness of motivational interviewing (MI) in reducing alcohol misuse for youth up to age 25. They concluded that the overall effect sizes of MI in this population were too small to be clinically meaningful. The present paper critically reviews the Foxcroft et al. meta-analysis, highlighting weaknesses, such as problems with search strategies, flawed screening and reviews of full-text articles, incorrect data abstraction and coding, and, accordingly, improper effect size estimation. In addition, between-study heterogeneity and complex data structures were not thoughtfully considered or handled using best practices for meta-analysis. These limitations undermine the reported estimates and broad conclusion made by Foxcroft et al. about the lack of MI effectiveness for youth. We call for new evidence on this question from better-executed studies by independent researchers. Meta-analysis has many important utilities for translational research. When implemented well, the overall effectiveness as well as different effectiveness for different populations can be examined via meta-analysis. Emerging methods utilizing individual participant-level data, such as integrative data analysis, may be particularly helpful for identifying the sources of clinical and methodological heterogeneity that matter. The need to better understand the mechanisms of alcohol interventions has never been louder in the addiction field. Through more concerted efforts throughout all phases of generating evidence, we may achieve large-scale evidence that is efficient and robust and provides critical answers for the field.

Meta-analysis is a quantitative synthesis methodology aimed at systematically searching and screening relevant studies and integrating the findings from the eligible single studies to generate more robust and efficient estimates. Meta-analysis shows the magnitude and consistency of the overall effect size under investigation across studies. When the effect is consistent, it can then be generalized to a larger population. Reflecting this important utility, there has been a meteoric rise in the number of meta-analysis publications (Sutton & Higgins, 2008). Furthermore, meta-analysis is increasingly seen as a method to combat the

current reproducibility crisis in preclinical and clinical trial research (Begley & Ellis, 2012; Ioannidis, 2005) and, more generally, as an important tool for open research practices (Cooper & VandenBos, 2013; Cumming, 2014; Eich, 2014; Open Science Collaboration, 2012). For these reasons, findings from meta-analyses weigh heavily in evidence-based decision making.

Meta-analysis studies, however, are not without limitations, some of which may be inherent to a method that relies on aggregated data (AD) from published studies (Cooper & Patell, 2009). In addition, like all methodologies, meta-analysis must be applied properly to yield valid conclusions. The present review was motivated by a recent meta-analysis by Foxcroft, Coombes, Wood, Allen, and Almeida Santimano (2014) for the Cochrane Database of Systematic Reviews that evaluated the effectiveness of motivational interviewing (MI) interventions for heavy drinking among adolescents and young adults up to age 25. Their study included 66 randomized controlled trials of individuals between the ages of 15 and 25, of which 55 studies were quantitatively analyzed for alcohol outcomes. Foxcroft et al. concluded that:

> There are no substantive, meaningful benefits of MI interventions for the prevention of alcohol misuse. Although some significant effects were found, we interpret the effect sizes as being too small, given the measurement scales used in the studies included in the review, to be of relevance to policy or practice (p. 2).

Foxcoft et al. further stated that the quality of evidence was low to moderate and, consequently, the reported effect sizes in their meta-analysis may be overestimated. The resulting press coverage declared "Counseling Has Limited Benefit on Young People Drinking Alcohol" (ScienceDaily news, 2014, August) and "Counseling Does Little to Deter Youth Drinking, Review Finds" (HealthDay news, 2014, August).

Given the potential impact that Foxcroft et al. (2014) may have on intervention and policy development, as well as clients' understanding of their treatment options, it is important to critically examine the study's methodology and conclusion. The quality of evidence is just as important for meta-analysis reviews as for single studies when making recommendations to clinicians in the field (see the Grading of Recommendations Assessment, Development and Evaluation [GRADE] approach; Guyatt et al., 2008). In discussing Foxcroft et al. for the present article, we utilize available reporting guidelines for meta-analysis studies, such as the *P*referred *R*eporting *I*tems for *S*ystematic reviews and *M*eta-*A*nalysis (PRISMA; Moher, Liberati, Tetzlaff, Altman, & The PRISMA group, 2009) of randomized controlled trials. The PRISMA has been adopted by many journals and organizations, including the Cochrane Collaboration, all *BMC* journals, and *PloS One*. The *M*eta-*A*nalysis *R*eporting *S*tandards (MARS) has been adopted by the American Psychological Association (APA) as part of the *J*ournal *A*rticle *R*eporting *S*tandards (JARS) by the APA Publications and Communications Board Working Group (2008). A new open access journal *Archives of Scientific Psychology* by APA (Cooper & VandenBos, 2013) further elaborates the MARS to help authors to improve the reporting of systematic reviews and meta-analyses. We also utilized other tutorial articles and textbooks (e.g., Borenstein, Hedges, Higgins, & Rothstein, 2009).

Review of the Foxcroft et al. (2014) Meta-analysis

Motivational interviewing (Miller & Rollnick, 2013) has been widely adopted for adolescents and young adults. Brief motivational interventions (BMIs) that are derived from MI, in particular, have become a popular choice among college administrators nationwide: 62% of schools utilizing an empirically-supported alcohol prevention program use BMIs (Nelson, Toomey, Lenk, Erickson, & Winters, 2010). Thus, the overall conclusion by Foxcroft et al. (2014), in which a majority of the data were from college BMIs, about the lack of clinically meaningful effects, appears controversial. Broadly speaking, the conclusion from Foxcroft et al. is inconsistent with those from other systematic reviews on college students (Cronce & Larimer, 2011; Larimer & Cronce, 2002; Larimer & Cronce, 2007) and meta-analysis studies (Carey, Scott-Sheldon, Carey, & DeMartini, 2007). In contrast to Foxcroft et al., these studies have generally supported the efficacy of MI-based interventions, despite inconsistencies across outcome measures, studies, and follow-up periods. In the sections below, we discuss our concerns about Foxcroft et al. in greater detail (see also Table 1 for a list).

Research Questions and Meta-analysis Methods

Foxcroft et al. (2014) had two research questions: (a) to examine the overall effect of MI for problematic drinking among youth up to age 25; and (b) to model the variability in effect sizes by the length of MI. Since the length of MI may be considered as exposure to treatment, the two aims were closely linked to provide an answer as to MI effectiveness in reducing drinking and related harm among youth. The included studies varied widely, including variation within the four major domains of populations identified by Cronbach (1982): unit of analysis (U; e.g., participants), treatments (T), outcome measures (O), and settings (S) in which a treatment takes place.

Given that meta-analysis is aimed at providing the overall inference across domains, one may argue that the broad inclusion of heterogeneous studies in a meta-analysis may make sense. However, this argument can be supported only when there are no important study-to-study variations (i.e., study-level variables) that may "systematically" affect the overall treatment effect size estimate. Combining effect sizes while ignoring their inconsistency or variability across studies in a meta-analysis would be conceptually analogous to interpreting main effects in the presence of an interaction effect in an analysis of variance, which is often not meaningful. Furthermore, in the psychological literature, variation in measures (e.g., reliability of measures and range of restriction) across effect sizes contribute considerably to overall variability, leading to a declaration that effect sizes "derived from dissimilar measures should rarely, if ever, be pooled" (Hedges & Olkin, 1985, p. 211).

Clinical and methodological heterogeneity vs. statistical heterogeneity—To discuss the issue of heterogeneity further, it is important to distinguish clinical and methodological heterogeneity from statistical heterogeneity. The presence of substantial statistical heterogeneity suggests that important between-study differences in the sample, measure, or treatment exist, while the reverse is not true. Statistical heterogeneity refers to the extent to which the variation in effect sizes across studies is above that expected by

chance, and is quantified and formally tested as part of a meta-analysis. In contrast, clinical and methodological heterogeneity more broadly refers to diversity across studies included in meta-analysis has been criticized as the "combining-apples-and-oranges" problem (see Bangert-Drowns, 1986). Whether it makes sense to combine apples and oranges is a judgment call that depends on the broader context of the research questions. If one were to make a fruit salad, it may make sense to combine apples and oranges, as Robert Rosenthal famously said (Borenstein et al., 2009, p. 357). It would not make sense, however, to combines apples and oranges if one's goal is to make an apple pie.

In Foxcroft et al. (2014), included studies differed in terms of sample demographic features (e.g., inclusion criteria, college vs. pre-college adolescents), clinical intervention characteristics (e.g., content, exposure, primary vs. secondary intervention target), and other study designs (e.g., outcome measures, follow-up schedules, comparison group, study quality). Some of the included MI interventions in the Foxcroft et al. meta-analysis targeted HIV and other risk behaviors (Naar-King et al., 2008), drug use (Marsden et al., 2006), alcohol-exposed pregnancy (Ceperich & Ingersoll, 2011), violence and alcohol misuse (Cunningham et al., 2012), as well as more alcohol-focused MI interventions. Consequently, the content of the MI and the manner of its implementation widely differed. Some studies did not have controls but instead used other bona fide interventions as their comparison group, while others had assessment only controls. By pooling effect sizes from such diverse studies, Foxcroft et al. may have undervalued important distinctions across populations.

As an example of potentially meaningful subgroups or treatment modifiers that were unexplored, the efficacy of MI was not examined separately for college vs. non-college or pre-college youth. Of the 66 studies reviewed, 37 trials were conducted on college campuses. Foxcroft et al. (2014) reasoned that "most studies were from college setting" and "this subgroup analysis would not have been meaningful or useful" (p. 140). However, it may not be appropriate to combine pre-college adolescents and non-college youth with college students when estimating the overall intervention effect size. First, evidence is clear that college students drink more heavily than their non-college peers (White & Hingson, 2013 for a review). In addition, although there are common risk factors for excessive alcohol use (e.g., high school drinking) between the two populations, college-specific risk factors do exist. In fact, there are many developmental and environmental factors specific to college environments that facilitate alcohol experimentation among college students (Schulenberg & Maggs, 2002).

Statistical heterogeneity and different meta-analysis models—To understand how between-study differences are modeled in meta-analysis, one must also appreciate the difference between fixed-effects and random-effects models, the two most common models for metaanalysis (see also Mun, Jiao, & Xie, in press for a review). Under the fixed-effects model, we assume there is one true effect size θ underlying all studies and thus any differences in observed effect sizes are due to sampling error s_i^2, which varies from study to study. Included studies are then considered as replications of the same trial that vary only due to sampling error. Studies are functionally interchangeable.

In contrast, under the random-effects model, we assume that there is no one true common effect size at the study level. Studies vary in true effect sizes depending on, for example, the characteristics of participants and study design. More formally, a study follows a distribution of effect size θ_i and within-study variability s_i^2 for i^{th} study, which is drawn from a superpopulation with mean θ and variance τ^2. Variance τ^2 represents study-to-study variation surrounding θ. Therefore, under the random-effects model, there are two sources of variation. They are within-study variance s_i^2 at the study level and between-study variance τ^2 at the superpopulation level. When $\tau^2 = 0$, results from a random-effects model become identical to those from a fixed-effects model. In the applied literature, there may a tendency to interpret a random-effects meta-analysis as if it were a fixed-effect model — that is, to treat the overall effect size from a random-effects model as if it were "the" true effect size. Another related misconception may be that the use of a random-effects model takes care of all between-study differences in effect sizes. In truth, the use of a random-effects model implies that effect sizes vary across studies by observed and unobserved features. Furthermore, the burden of correctly identifying the sources of the between-study differences is borne by the meta-analysts. There are several ways that variability in effect sizes can be gauged.

Measures of heterogeneity—To formally test whether effect sizes are homogeneous across studies, the Q statistic (Cochran, 1950) is calculated as below:

$$Q = \sum_i W_i (Y_i - \overline{Y}_w)^2,$$

where W_i is the inverse of the sampling variance for study i and $\overline{Y}_w = \sum_i W_i Y_i / \sum_i W_i$ is the weighted estimator of the effect of interest. Under the null hypothesis, Q follows a chi-squared distribution with $k-1$ degrees of freedom. If rejected, it signals that statistical heterogeneity is beyond chance. However, this Q statistic is overly sensitive to the number of studies included in meta-analysis. Thus, an alternative measure I^2 has been proposed (Higgins & Thompson, 2002) and used widely along with the Q statistic in applied meta-analyses:

$$I^2 = \frac{\hat{\tau}^2}{\hat{\tau} + \hat{\sigma}^2},$$

where $\hat{\sigma}^2 = s^2$. I^2 stands for the proportion of variance of effect sizes unexplained that can be attributed to between-study differences, and is reported in percentage. An I^2 value less than 30% may indicate mild heterogeneity; notable heterogeneity if substantially exceeding 50% (Higgins & Thompson, 2002).

In Foxcroft et al. (2014), both Q test and I^2 value were reported for all 16 outcome analyses. I^2 values exceeded 50% in seven outcome analyses: alcohol quantity and frequency, binge

drinking, and risky behavior at < 4 months follow-up, alcohol problems at < 4 months and 4+ months follow-ups, and drink driving at 4+ months follow-up. In addition, in their analysis, MI interventions were compared not only to assessment only controls but also to other bona fide interventions that are known to be efficacious, such as stand-alone normative feedback (Miller et al., 2013; Walters & Neighbors, 2005). Not surprisingly, there was considerable heterogeneity in the overall effect size estimates from two submodels, one of which compared MI interventions against alternative interventions; the other against assessment only controls. Out of the nine analyses that analyzed both submodels, three effect sizes differed considerably: alcohol quantity ($I^2 = 73\%$), frequency (85%), and binge drinking (56%) at < 4 months follow-up. In sum, the variability in effect sizes across the studies included in Foxcroft et al. (2014) was considerable especially for outcomes assessed at a short-term follow-up; and also between the two metaanalysis models that used different controls. Therefore, we can conclude that nonignorable heterogeneity was found in many outcome analyses, which raises a concern that highly heterogeneous studies were combined.

In conclusion, a study aimed at examining the overall effect size, such as Foxcroft et al. (2014), should pool data only from sufficiently similar studies when estimating the overall effect sizes, even if doing so may limit the scope of the meta-analysis and, as a consequence, result in the loss of some information. The inclusion of heterogeneous studies in meta-analysis might have been helpful, had there been further investigations on the reasons for different treatment effect sizes by including informative covariates in the model. Beyond that τ^2 estimates and I^2 values were reported in all analyses and Q tests were reported in Foxcroft et al., no further actions were taken to address between-study heterogeneity in effect sizes.

Study Inclusion and Exclusion

Deciding which studies to include or exclude in a meta-analysis is very important as it addresses the systematic, nonbiased, and comprehensive nature of a review. Figure 1 shows the recommended PRISMA flow diagram (Moher et al., 2009; www.prisma-statement.org). This flow diagram for meta-analysis is akin to a CONSORT flow diagram for a randomized controlled trial (RCT), which was developed to ensure transparent reporting of trials (Schulz, Altman, Moher, & The CONSORT Group, 2010). An accurate flowchart of studies through several key stages of screening and reviewing studies is important as it helps readers to understand meta-analysis study procedures and to gauge whether conclusions are valid.

In Foxcroft et al. (2014), we observed several weaknesses with respect to searching and documenting eligible studies. First, we note that multiple publications from the same trial were sometimes counted twice. For example, "White 2006" (White, Morgan, Pugh, Celinska, Labouvie, & Pandina, 2006) and "White 2007" (White, Mun, Pugh, & Morgan, 2007) are from the same trial. White et al. (2006) reported data at the first, short-term follow-up, whereas White et al. (2007) reported both short-term and long-term follow-up data. However, White et al. (2006) should have been removed at an earlier stage as a duplicate because it reports partial data. Similarly, "Walton 2010" (Walton et al., 2010) and "Cunningham 2012" (Cunningham et al., 2012) are from the same trial. Walton et al. reports data up to 6 month follow-up, whereas Cunningham et al. reports data at 12-month follow-

up. However, both "Walton 2010" and "Cunningham 2012" were included in the final tally of the 55 studies for quantitative meta-analysis, and subsequently analyzed together in analyses. This double counting of the same or similar data can arbitrarily inflate the sample size and underestimate standard errors, and may lead to liberal statistical decisions. The text of these articles from the same trials clearly notes this connection between studies, and hence their inclusion as independent studies in Foxcroft et al. raises a concern whether reviewers had sufficient content understanding.

Second, meta-analysis investigators also obtain data through individual contacts with study authors or well-known authors in the field. Unpublished studies are often obtained through peer consultation and by contacting specific investigators who are active in the field (Berman & Parker, 2002). Contacting active investigators might have been particularly helpful for Foxcroft et al. (2014) because in the field of BMIs for college students, each college campus adopts a unique name for their intervention, and hence it is possible that some of the MI interventions do not come up via electronic searches. To illustrate this point, the Lifestyle Management Class reported by Fromme and Corbin (2004) can be considered as an MI delivered in groups (Carey et al., 2007; Ray et al., 2014), which also provided personalized feedback. Similarly, Amaro et al. (2009) provided a BMI for mandated college students based on the Brief Alcohol Screening and Intervention for College Students program (Dimeff, Baer, Kivlahan, & Marlatt, 1999) but described their intervention as the University Assistance Program. If experts in the field were contacted or if the existing reviews (e.g., Cronce and Larimer, 2011) were cross-referenced, these studies would likely have been included. The omission of relevant studies in Foxcroft et al. may also be attributed to the fact that Foxcroft et al. screened studies based on the title and abstract rather than full-text articles.

Third, when there were multiple intervention conditions within studies, Foxcroft et al. (2014) counted these interventions multiple times as if they were independent. For example, the study by Walters, Vader, Harris, Field, and Jouriles (2009) was counted as two studies toward the total of 55 studies that were quantitatively analyzed. Walters et al. had multiple intervention conditions: MI with feedback, MI without feedback, and feedback only, which was used as an alternative control in Foxcroft et al. However, it is incorrect to treat these multiple comparisons from the same study as if they are from two different studies because this practice misrepresents the extent of the systematic review (i.e., inflating the number of studies); it also violates the assumption involved in subsequent analyses when combining estimates. Out of the final tally of 55 "studies," many "studies" were actually "subgroups" within studies, such as mandated vs. volunteer students or different treatment groups. Furthermore, multiple effect sizes (multiple related outcomes) from studies were counted independently. The handling of complex data structures related to multiple subgroups, interventions, and outcomes within studies as independent observations raises questions of nonindependence among estimates and inflated sample size. Overall, this practice affects an estimate of the precision surrounding the overall effect size (variability surrounding θ) in random-effects models. When some studies have multiple groups whereas the vast majority of studies do not, computing a combined treatment effect per study and making study as a unit of analysis may be an option (Borenstein et al., 2009). More generally, multiple related

parameters may be combined more appropriately by using multivariate metaanalysis that takes into account within-study correlations (Jackson, Riley, & White, 2011 for a review).

Fourth, we noted generally poor record keeping in Foxcroft et al. (2014). For example, we observed that three different studies by the same author were listed in the references section and/or mentioned in the main text of the meta-analysis: "Walters 2000", "Walters 2005", and "Walters 2009." However, "Walters 2005" is a review study (Walters & Neighbors, 2005) and should have been screened out and, consequently, excluded, prior to examining its eligibility. Instead, this review study was counted toward 30 studies whose full-text articles were assessed for eligibility and subsequently excluded with reasons (i.e., ineligible intervention, age of participants, non-relevant outcome measures, and study design). Furthermore, "Walters 2005" was referenced incorrectly as Walters, Bennett, and Miller (2000) on p. 34 in Foxcroft et al. The mishandling of these two very different studies suggests that there may have been problems in screening, reviewing and/or tallying studies across key stages of the review.

Finally, Foxcroft et al. (2014) reportedly used a standardized data extraction form that was not made available. More critically, the expertise and training background of reviewers was not revealed; screening involved examining the titles and abstracts of studies without looking at full-text articles; and the extent of disagreement among reviewers was not documented. Information about the qualifications of reviewers (in both expertise and training) is recommended for reporting on the checklist for publication of meta-analysis studies in *Archives of Scientific Psychology* by the APA (Cooper & VandenBos, 2013). It is also recommended that deidentified papers be evaluated by at least two reviewers, one of whom is a content expert who is knowledgeable in the subject matter and the other a biostatistician or epidemiologist who can evaluate the analytic methods (Berman & Parker, 2002). These observations may help to explain the weaknesses discussed in this section. In sum, the strategies for searching, screening, and reviewing relevant articles in Foxcroft et al. were neither comprehensive nor thorough, a serious concern for any meta-analysis.

Data Abstraction and Coding

After identifying appropriate studies to be included in a meta-analysis, the next step is to correctly abstract and code the data. Several inaccuracies we noted are (a) inaccurate data pooling with respect to follow-up periods; (b) inaccurate data coding and effect size estimates; and (c) no correction for small samples.

Foxcroft et al. (2014) pooled outcome data according to the length of the time since the intervention, using a binary cut at less than 4 months, or 4+ months in the meta-analysis review. For example, Walters et al. (2009) had short-term outcomes at 3-mo. follow-up, which were omitted in the short-term (< 4 months) analysis. In contrast, data from Martens, Smith, and Murphy (2013) was used for both short-term and long-term analyses (4+ months). For Walters et al. (2000) which followed-up participants at 6 weeks, Foxcroft et al. (2014) included their data in the analysis of alcohol-related problems at 4+ months follow-up, rather than in the analysis of problems at < 4 months follow-up. Thus, data from eligible studies seem to have been excluded in subsequent analyses, and some of the included data were, in actuality, ineligible for synthesis.

We also note errors in data coding and, consequently, incorrect effect size estimates. For example, Walters et al. (2009) was a dismantling study that utilized a 2×2 factorial design to examine the effects of MI with, and without, feedback, as well as the effects of feedback alone without MI. Therefore, there were two MI groups relevant for Foxcroft et al.'s study (2014): MI with feedback (MIF) and MI without feedback (MIO). These two MI groups were compared against the feedback only (FBO) and assessment only (AO) controls, respectively, in the metaanalysis. Their calculation showed a favorable alcohol quantity outcome toward FBO and, consequently, a negative outcome for MIF. The reported standardized mean difference (SMD) estimate in the meta-analysis was 0.07 (SE = 0.18). Our calculation (see also Table 2), however, shows that the SMD should have been –0.18 (SE = 0.18), favoring the MIF group instead. This error represents a difference of 0.25 in magnitude and *opposite* in direction to the effect reported in the original study. In addition, for the comparison between MIO and assessment only control (AO), the meta-analysis review estimated the SMD to be –0.05 with SE = 0.18. However, the correct SMD is –0.11 (SE = 0.18). Given that the reported SMDs and SEs were the same for both alcohol quantity and alcohol problems in Foxcroft et al., it is possible that the means and standard deviations for alcohol problems may have been incorrectly entered to a database to calculate SMDs for both alcohol quantity and alcohol problems outcomes. Below, we describe SMD and SE, and how they can be calculated.

Computing *SMD* and *SE* estimates—When examining each published study reported in Foxcroft et al. (2014), the vast majority of the studies reported a full set of summary data, including the mean, standard deviation, and sample size for each group. In that case, the computation of the effect size and its variance is straightforward. The SMD can be obtained by dividing the mean difference in each study by that study's standard deviation, which would then be comparable across studies that used different outcome measures.

For instance, in studies that used two independent groups (e.g., treatment and control), we can estimate the SMD as follows:

$$d = \frac{\overline{x}_t - \overline{x}_c}{s_{within}},$$

where the treatment group is indicated by subscript *t* and the control group by subscript *c*. The pooled standard deviation of the two groups can be calculated

$$s_{within} = \frac{\sqrt{(n_t - 1)s_t^2 + (n_c - 1)s_c^2}}{n_t + n_c - 2},$$

where *s* indicates standard deviation and *n* indicates sample size. The variance of *d* is then

computed by $V_d = \frac{n_t + n_c}{n_t n_c} + \frac{d^2}{2(n_t + n_c)}$ and the standard error of *d* is the square root of the

variance of d. In small samples, d is biased and an unbiased version called Hedges' g (Hedges, 1981) is typically recommended. Hedges' g is computed:

$$g = J \times d, \text{ with } J = 1 - \frac{3}{4\mathrm{d}f - 1} \text{ and } SE_g = \sqrt{J^2 \times V_d}.$$

In the equation above, df is the degrees of freedom used to estimate s_{within}. For two independent groups, this would be $n_t + n_c - 2$. The correction factor J (an approximation in the equation above) will always be less than 1 in small studies. When J is 1, the corrected *SMD* estimate g is the same as d. For a two-arm trial with a total of 100 participants that were randomly assigned, J is 0.99, which means that d from a small trial of $N = 100$ will be reduced by 1% when correcting for small samples. For a two-arm trial with a total of 50 participants, J becomes 0.98. Consequently, g estimates will always be less than d and the variance of g will always be less than variance of d.

In the event that there is no control group in within-subject design studies (pre-post), *SMD* estimates can be calculated as follows:

$$d = \frac{\overline{x}_b - \overline{x}_f}{s_{within}},$$

where the baseline score is indicated by subscript b and the follow-up score by subscript f. The standard deviation within groups can be obtained from the standard deviation of the difference by using

$$s_{within} = \frac{s_{diff}}{\sqrt{2(1 - r)}},$$

where r stands for the correlation between baseline and follow-up measures. The variance of d is then computed by $V_d = 2(1 - r)((1/n) + (d^2/2n))$ and the standard error of d is the square root of the variance of d (namely, V_d). For small samples, Hedges' g can also be calculated, with the single exception that $J = n - 1$ for related samples. For studies that report log odds ratio, the log odds ratio can be converted to d by multiplying the log odds ratio by $\frac{\sqrt{3}}{\pi}$ where $\pi = 3.14$. The variance of d would then be the variance of the log odds ratio multiplied by $\frac{3}{\pi^2}$ (Borenstein et al., 2009). In theory, d or Hedges' g can be combined across studies that use different designs (i.e., RCT or within-subject designs) or analyses (i.e., continuous or binary outcomes), although different designs or analyses may introduce other sources of heterogeneity.

Foxcroft et al. (2014) did not make adjustments for d for small samples. Although this correction may not have dramatically affected the overall estimates, it is typical in the field

to adjust for the effect sizes of small samples (e.g., Carey et al., 2007; Scott-Sheldon, Carey, Elliott, Garey, & Carey, 2014).

Recalculation of *SMD* estimates—Using the equations shown above, we examined published data for two outcome measures—alcohol quantity and alcohol-related problems—for 4+ months follow-up to gauge whether the reported miscalculation of *SMD* were isolated incidents and, if not, how frequently it occurred.

Of the 28 studies included in alcohol quantity outcome analysis in the Foxcroft et al. (2014) review, our recalculation showed that two changed the direction of effect size estimates: Marlatt et al. (1998; reported *SMD* = 0.14 vs. our calculation = –0.15) and Walters et al. (2009; MIF vs. FBO; reported *SMD* = 0.07 vs. our calculation = –0.18). In addition, two studies changed their effect size magnitude considerably: McCambridge and Strang (2004; reported *SMD* = –0.08 vs. our calculation = –0.25) and Walters et al. (2009; MIO vs. controls; reported *SMD* = –0.05 vs. our calculation = –0.11). Similarly, of the 24 studies included in Foxcroft et al. for the analysis of alcohol problems, our recalculation showed that two studies changed the direction *of SMDs:* Larimer et al. (2001; reported *SMD* = –0.09 vs. our calculation = 0.09) and Terlecki (2011; indicated as "Terlecki, 2010a" volunteer students; reported *SMD* = 0.11 vs. our calculation = –0.29). In addition, for Cunningham et al. (2012), the direction of *SMDs* differed depending on the source of the data (i.e., 2 × 2 outcome data or generalized estimating equation analysis, which treat missing data differently). Note also that sample size for studies was inconsistently coded in Foxcroft et al., which, consequently, affects *SMD* and *SE* calculation. In addition, we observed when there are multiple follow-ups, effect sizes from the longest follow-up were systematically chosen in Foxcroft et al. Given that intervention effects tend to diminish over time, this strategy could have resulted in a downward bias toward overall treatment effect size estimates.

Our estimates are consistent with the published data and the stated conclusion of the original studies. Therefore, we cautiously speculate that the *SMD* estimates reported in Foxcroft et al. (2014) that are different from our estimates might have been caused by the errors in data abstraction and/or coding, which, on average, would have led to underestimated effect sizes.

Data Analysis, Data Presentation, and Interpretation

Estimated *SMDs* may be technically correct but can be invalid at the same time. For example, Marsden et al. (2006) randomly assigned stimulant users to either an intervention or a control condition. This study reported no between-subject, intervention effects on any of the outcomes, alcohol use included, at 6-month follow-up. Overall, there was very little change in substance use, and change in alcohol use was not statistically significant. Hence, Marsden et al. concluded that "our brief motivational intervention was no more effective at inducing behaviour change than the provision of information alone." (p. 1014). However, in the analysis by Foxcroft et al. (2014), this was one of the few studies that showed a statistically significant, strong intervention effect on alcohol use, with *SMD* = –0.35 (*SE* = 0.11).

Figure 2 explains this paradox between the strong intervention effect shown in Foxcroft et al. (2014) and the null finding in Marsden et al. (2006). Figure 2 was drawn using the data reported in Table 3 of Marsden et al. First, Figure 2a shows a statistically significant, between-group difference in alcohol use in favor of the MI intervention group at 6-month follow-up. However, when we include baseline data for these two groups, it is clear that there were very little changes from baseline to follow-up in both groups, and that there were no discernible between-group differences in change in alcohol use (see Figure 2b). Marsden et al. discussed disappointing results for alcohol consumption and attributed the results to the fact that drinking was not the major focus of the intervention, and that there were baseline differences. Therefore, Marsden et al. clearly presented their data and discussed their null findings.

SMD can be estimated from data showing unadjusted mean differences, adjusted mean differences, log odds ratios, and change scores from pre-post designs. Thus, extracting data from published studies while not taking into account important features of each study can result in nonsensical estimates, as was the case in the handling of the published data in Marsden et al. (2006) in the Foxcroft et al. (2014) review. This observation echoes our concerns about whether reviewers had sufficient content understanding and whether reviews were done thoroughly.

Given these Shortcoming, is the Conclusion of Foxcroft et al. (2014) Wrong?

The simple answer is we do not know. Scientific conclusions are valid only when the methods are valid. Can we be confident that the estimates and conclusion are correct? Here, we would firmly answer no. Given the nature of the concerns and weaknesses detailed in the present study, we are very uncertain about the reported estimates in Foxcroft et al. (2014). The weaknesses we observed could have affected the estimates in opposite directions and, consequently, we may not know the correct estimates and the validity of the overall conclusion without new investigations. Thus, it may be best to reserve our judgment about their broad conclusion regarding the efficacy of MI for youth until we have evidence from more carefully executed reviews by other independent researchers. Alcohol consumption remains a leading preventable cause of death among college students (Hingson, Zha, & Weitzman, 2009) and has been relatively stable since 1976 when NIAAA first reported on it (White & Hingson, 2013). Up to 63% of young people report drinking, and 35% report heavy episodic drinking in the past month (Johnston, O'Malley, Bachman, & Schulenberg, 2013). Therefore, it is critical to identify evidence-based practices for reducing drinking and mitigating risk among youth.

Discussion

Scientific investigations rely on the self-correcting nature of the peer-review process and ensuing discussions within the greater research community following publication. It is important to critically and timely examine findings from all empirical investigations, including meta-analyses, to find areas for improvement and to channel our efforts for better directions in future research. Such efforts are critical for understanding mechanisms of change, a critical step towards evidence-based treatment and, ultimately, the Precision

Medicine Initiative introduced by President Obama during the State of the Union Address in 2015 (The White House, 2015).

The need to improve our understanding of why certain "brand treatments" and "treatment components" work "for whom" has been central; yet progress has been slow in the addiction field (Miller & Moyers, 2015). Miller and Moyers pointed out that whereas a handful of positive RCTs may be sufficient to demonstrate the efficacy of a new medication, there are more than 400 RCTs published for treatments of alcohol use disorders alone. To put these numbers in another context, the median number of studies that are reviewed in the Cochrane Database of Systematic Reviews is six (Davey, Turner, Clarke, & Higgins, 2011). In contrast, despite the plethora of data in the addiction field, it is unclear, using the expression used by Miller and Moyers, which horse won the race and what led to the top performance. To fully understand the mechanisms of successful treatments, Miller and Moyers argued for the need to look more closely at general or common treatment factors (e.g., relationship features between therapists and clients) instead of limiting our attention exclusively to treatment-specific factors. Similarly, Gaume, McCambridge, Bertholet, and Daeppen (2014) pointed out in their recent narrative review that the term "brief intervention" refers to a group of heterogeneous interventions conducted in different settings, which has been treated like a black box, when its content has been substantially modified over the years and across studies without much efforts to identify what contents of the black box work at the mechanistic level.

A quantitative research synthesis of large-scale data pooled from multiple sources, when appropriately designed and executed, can help answer these questions about the individual, relational, and contextual conditions under which interventions work well. A newly emerging meta-analysis approach using individual participant-level data (IPD) (Cooper & Patall, 2009), which is how it is generally called in the field of medical research but is referred to as integrative data analysis in psychological research (IDA; Curran & Hussong, 2009; Hussong, Curran, Bauer, 2013), may nicely complement traditional AD meta-analysis: The IPD meta-analysis approach is well suited to directly tackle the sources of the clinical and methodological heterogeneity and to explore important subgroups of participants, interventions, settings, and follow-ups that show different intervention effects. While AD meta-analysis can be much more quickly conducted for a larger number of studies, compared to IPD meta-analysis, the latter may prove to be critical for obtaining robust and efficient estimates for the field where AD meta-analysis may fall short.

In one of the first research synthesis projects that utilized IPD pooled from BMIs for college students, Mun et al. (2015) indicated that the operational definitions of "treatment" and "control" differed across studies. Therefore, there was a need to harmonize these intervention labels across studies to be able to meaningfully compare apples with apples, and oranges with oranges. Harmonization of treatment groups involved detailed coding of all intervention content materials as well as their delivery characteristics by multiple content experts. There was substantial variability in the number of topics and the content nature targeted by BMIs across studies, which affected intervention outcomes (see Ray et al., 2014). To examine intervention effects, outcome measures can be harmonized and/or individual trait scores may be derived using advanced measurement models (see Huo et al.,

2014 for technical details). We analyzed IPD from 38 intervention groups (i.e., Individual MI with Personalized Feedback [PF], PF only, Group MI, and controls) from 17 studies and found that college BMIs did not have statically significant, overall intervention effects on alcohol use quantity, peak drinking, and alcohol problems, after adjusting for baseline individual demographic characteristics and drinking levels (Huh et al., 2015). The findings suggest that BMIs, on average, may have little to no effects, and that there is a need for better overall intervention strategies for college students, as well as for specific strategies meeting different needs for different college populations. Research synthesis studies that combine data more efficiently using all available data (both AD and IPD), and studies that directly examine different responses to interventions via subgroup analysis or equivalent analysis may provide better answers to the questions raised in the present study.

Conclusion

In this critical review, we recommend that the "vote of no confidence" in MI for youth by Foxcroft et al. (2014) be rejected until we have new evidence from better-executed reviews by independent researchers. Meta-analysis is not above scrutiny; just as single studies are scrutinized by independent researchers. Therefore, there is a need to critically examine findings from meta-analysis reviews. When a meta-analysis becomes available, authors of included studies or potentially eligible studies may want to check to see whether their work was accurately included or excluded, and whether their data were correctly represented in the meta-analysis. In addition, it will be helpful for clinical trial researchers to provide all necessary information more clearly so that their trials are easily searchable and data can be extracted without causing challenges or errors for meta-analysis or any other research synthesis studies. When the quality and transparency of individual trials are improved, it will be much easier to conduct research synthesis studies (see also Mun et al., 2015). In return, we can have more confidence in our findings from the upstream to downstream of the evidence generating processes. Through more concerted systematic efforts and communications from all parties within the addiction research community, we can establish more rigorous and transparent research practice that has been called for by the National Institutes of Health (Collins & Tabak, 2014) and major journals in the field (e.g., Eich, 2014). Ultimately, the improved efforts in all phases of scientific inquiries can get us closer to more personalized treatment decisions, a challenge for the current generation of translational researchers.

Acknowledgments

References

Amaro H, Ahl M, Matsumoto A, Prado G, Mulé C, Kemmemer A, et al. Mantella P. Trial of the University Assistance Program for alcohol use among mandated students. Journal of Studies on Alcohol and Drugs, Supplement. 2009; 16:45–56. [PubMed: 19538912]

American Psychological Association Publications and Communications Board Working Group on Journal Article Reporting Standards. Reporting standards for research in psychology: Why do we need them? What might they be? 2008. American Psychologist. 2008; 63:839–851.10.1037/0003-066x.63.9.839 [PubMed: 19086746]

Bangert-Drowns RL. Review of developments in meta-analytic method. Psychological Bulletin. 1986; 99(3):388–399.10.1037/0033-2909.99.3.388

Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature. 2012; 483(7391):531–533.10.1038/483531a [PubMed: 22460880]

Berman N, Parker R. Meta-analysis: Neither quick nor easy. BMC Medical Research Methodology. 2002; 2(1):10.10.1186/1471-2288-2-10 [PubMed: 12171604]

Borenstein, M.; Hedges, LV.; Higgins, JPT.; Rothstein, HR. Introduction to meta-analysis. New York, NY: Wiley; 2009.

Carey KB, Scott-Sheldon LA, Carey MP, DeMartini KS. Individual-level interventions to reduce college student drinking: a meta-analytic review. Addictive Behaviors. 2007; 32(11):2469–2494.10.1016/j.addbeh.2007.05.004 [PubMed: 17590277]

Ceperich S, Ingersoll K. Motivational interviewing + feedback intervention to reduce alcohol-exposed pregnancy risk among college binge drinkers: Determinants and patterns of response. Journal of Behavioral Medicine. 2011; 34(5):381–395.10.1007/s10865-010-9308-2 [PubMed: 21318412]

Cochran WG. The comparison of percentages in matched samples. Biometrika. 1950; 37(3/4):256–266.10.1093/biomet/37.3-4.256 [PubMed: 14801052]

Collins FS, Tabak LA. NIH plans to ehnance reproducibility. Nature. 2014; 505:612–213. [PubMed: 24482835]

Cooper H, Patall EA. The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. Psychological Methods. 2009; 14(2):165–176.10.1037/a0015565 [PubMed: 19485627]

Cooper H, VandenBos GR. Archives of scientific psychology: A new journal for a new era. Archives of Scientific Psychology. 2013; 1(1):1–6.10.1037/arc0000001

Cronbach, LJ. Designining evaluations of educational and social programs. San Francisco, CA: Jossey Bass; 1982.

Cronce JM, Larimer ME. Individual-focused approaches to the prevention of College student drinking. Alcohol Research & Health. 2011; 34:210–221. [PubMed: 22330220]

Cumming G. The new statistics: Why and how. Psychological Science. 2014; 25(1):7–29.10.1177/0956797613504966 [PubMed: 24220629]

Cunningham RM, Chermack ST, Zimmerman MA, Shope JT, Bingham CR, Blow FC, Walton MA. Brief motivational interviewing intervention for peer violence and alcohol use in teens: One-year follow-up. Pediatrics. 2012; 129(6):1083–1090.10.1542/peds.2011-3419 [PubMed: 22614776]

Curran PJ, Hussong AM. Integrative data analysis: The simultaneous analysis of multiple data sets. Psychological Methods. 2009; 14(2):81–100.10.1037/a0015914 [PubMed: 19485623]

Davey J, Turner R, Clarke M, Higgins J. Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: A cross-sectional, descriptive analysis. BMC Medical Research Methodology. 2011; 11(1):160.10.1186/1471-2288-11-160 [PubMed: 22114982]

Dimeff, LA.; Baer, JS.; Kivlahan, DR.; Marlatt, GA. Brief alcohol screening and intervention for college students. New York, NY: Guilford; 1999.

Eich E. Business not as usual. Psychological Science. 2014; 25(1):3–6.10.1177/0956797613512465 [PubMed: 24285431]

Foxcroft DR, Coombes L, Wood S, Allen D, Almeida Santimano NML. Motivational interviewing for alcohol misuse in young adults. Cochrane Database of Systematic Reviews. 2014; (8) Art. No.: CD007025. 10.1002/14651858.CD007025.pub2

Fromme K, Corbin W. Prevention of heavy drinking and associated negative consequences among mandated and voluntary college students. Journal of Consulting and Clinical Psychology. 2004; 72(6):1038–1049.10.1037/0022-006X.72.6.1038 [PubMed: 15612850]

Gaume J, McCambridge J, Bertholet N, Daeppen JB. Mechanisms of action of brief alcohol interventions remain largely unknown – A narrative review. Frontiers in Psychiatry. 2014; 510.3389/fpsyt.2014.00108

Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ. What is "quality of evidence" and why is it important to clinicians? BMJ. 2008; 336(7651):995–998.10.1136/bmj. 39490.551019.BE [PubMed: 18456631]

HealthDay. Counseling does little to deter youth drinking, review finds. 2014 Aug. Retrieved from http://consumer.healthday.com/kids-health-information-23/kids-and-alcohol-health-news-11/counseling-does-little-to-deter-youth-drinking-review-finds-690892.html

Hedges L. Distribution theory for Glass's estimator of effect size and related estimators. Journal of Educational Statistics. 1981; 6:107–128.

Hedges, L.; Olkin, I. Statistical methods for meta-analysis. New York: Academic Press; 1985.

Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Statistics in Medicine. 2002; 21(11):1539–1558.10.1002/sim.1186 [PubMed: 12111919]

Hingson RW, Zha W, Weitzman ER. Magnitude of and trends in alcohol-related mortality and morbidity among U.S. college students ages 18-24, 1998-2005. Journal of Studies on Alcohol and Drugs. 2009; Suppl(16):12–20. [PubMed: 19538908]

Huh D, Mun EY, Larimer ME, White HR, Ray AE, Rhew I, et al. Atkins DC. Brief motivational interventions for college student drinking may not be as powerful as we think: An individual participant-level data meta-analysis. Alcoholism, Clinical and Experimental Research. 2015; 39:919–931.

Huo Y, de la Torre J, Mun EY, Kim SY, Ray AE, Jiao Y, White HR. A hierarchical multi-unidimensional IRT approach for analyzing sparse, multi-group data for integrative data analysis. Psychometrika. 201410.1007/s11336-014-9420-2

Hussong AM, Curran PJ, Bauer DJ. Integrative data analysis in clinical psychology research. The Annual Review of Clinical Psychology. 2013; 9:61–89.10.1146/annurev-clinpsy-050212-185522

Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005; 2(8):e124.10.1371/journal.pmed.0020124 [PubMed: 16060722]

Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. Statistics in Medicine. 2011; 30(20):2481–2498.10.1002/sim.4172 [PubMed: 21268052]

Johnston, LD.; O'Malley, PM.; Bachman, JG.; Schulenberg, JE. Monitoring the Future national survey results on drug use, 1975-2012 Volume II: College students and adults ages 19-50. Ann Arbor, MI: Institute for Social Research, The University of Michigan; 2013.

Larimer ME, Cronce JM. Identification, prevention and treatment: A review of individual-focused strategies to reduce problematic alcohol consumption by college students. Journal of Studies on Alcohol, Supplement. 2002; 14:148–163. [PubMed: 12022721]

Larimer ME, Cronce JM. Identification, prevention, and treatment revisited: individual-focused college drinking prevention strategies 1999-2006. Addictive Behaviors. 2007; 32:2439–2468.10.1016/j.addbeh.2007.05.006 [PubMed: 17604915]

Larimer ME, Turner AP, Anderson BK, Fader JS, Kilmer JR, Palmer RS, Cronce JM. Evaluating a brief alcohol intervention with fraternities. Journal of Studies on Alcohol. 2001; 62:370–380. [PubMed: 11414347]

Marlatt GA, Baer JS, Kivlahan DR, Dimeff LA, Larimer ME, Quigley LA, Williams E. Screening and brief intervention for high-risk college student drinkers: Results from a 2-year follow-up assessment. Journal of Consulting and Clinical Psychology. 1998; 66(4):604–615.10.1037/0022-006x.66.4.604 [PubMed: 9735576]

Marsden J, Stillwell G, Barlow H, Boys A, Taylor C, Hunt N, Farrell M. An evaluation of a brief motivational intervention among young ecstasy and cocaine users: no effect on substance and alcohol use outcomes. Addiction. 2006; 101(7):1014–1026.10.1111/j.1360-0443.2006.01290.x [PubMed: 16771893]

Martens MP, Smith AE, Murphy JG. The efficacy of single-component brief motivational interventions among at-risk college drinkers. Journal of Consulting and Clinical Psychology. 2013; 81(4):691–701.10.1037/a0032235 [PubMed: 23506464]

McCambridge J, Strang J. The efficacy of single-session motivational interviewing in reducing drug consumption and perceptions of drug-related risk and harm among young people: results from a multi-site cluster randomized trial. Addiction. 2004; 99(1):39–52.10.1111/j.1360-0443.2004.00564.x [PubMed: 14678061]

Miller WR, Moyers TB. The forest and the trees: relational and specific factors in addiction treatment. Addiction. 2015; 110(3):401–413.10.1111/add.12693 [PubMed: 25066309]

Miller, WR.; Rollnick, S. Motivational interviewing: Helping people change. New York, NY: Gilford; 2013.

Miller MB, Leffingwell T, Claborn K, Meier E, Walters S, Neighbors C. Personalized feedback interventions for college alcohol misuse: An update of Walters & Neighbors 2005. Psychology of Addictive Behaviors. 2013; 27(4):909–920.10.1037/a0031174 [PubMed: 23276309]

Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med. 2009; 6(7):e1000097.10.1371/journal.pmed.1000097 [PubMed: 19621072]

Mun EY, de la Torre J, Atkins DC, White HR, Ray AE, Kim SY, et al. The Project INTEGRATE Team. Project INTEGRATE:An integrative study of brief alcohol intervention trials for college students. Psychology of Addictive Behavior. 2015; 29:34–48.10.1037/adb0000047

Mun, EY.; Jiao, Y.; Xie, M. Integrative data analysis for research in developmental psychopathology. In: Cicchetti, D., editor. Developmental psychopathology. 3rd. Hoboken, NJ: Wiley; in press

Naar-King S, Lam P, Wang B, Wright K, Parsons JT, Frey MA. Brief Report: Maintenance of effects of motivational enhancement therapy to improve risk behaviors and HIV-related health in a randomized controlled trial of youth living with HIV. Journal of Pediatric Psychology. 2008; 33(4):441–445.10.1093/jpepsy/jsm087 [PubMed: 17905800]

Nelson TF, Toomey TL, Lenk KM, Erickson DJ, Winters KC. Implementation of NIAAA College Drinking Task Force recommendations: How are colleges doing 6 years later? Alcoholism: Clinical and Experimental Research. 2010; 34:1687–1693.10.1111/j.1530-0277.2010.01268.x

Open Science Collaboration. An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. Perspectives on Psychological Science. 2012; 7(6):657–660.10.1177/1745691612462588 [PubMed: 26168127]

Ray AE, Kim SY, White HR, Larimer ME, Mun EY, Clarke N. The Project INTEGRATE Team. When less is more and more is less in brief motivational interventions: Characteristics of intervention content and their associations with drinking outcomes. Psychology of Addictive Behaviors. 2014; 28:1026–1040.10.1037/a0036593 [PubMed: 24841183]

Schulenberg JE, Maggs JL. A developmental perspective on alcohol use and heavy drinking during adolescence and the transition to young adulthood. Journal of Studies on Alcohol. 2002; Supplement(14):54–70.

Science Daily. Counseling has limited benefit on young people drinking alcohol, study suggests. 2014 Aug. Retrieved from http://www.sciencedaily.com/releases/2014/08/140821090653.htm

Schulz KF, Altman DG, Moher D. The CONSORT Group. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. PLoS Med. 2010; 7(3):e1000251.10.1371/journal.pmed.1000251 [PubMed: 20352064]

Scott-Sheldon LAJ, Carey KB, Elliott JC, Garey L, Carey MP. Efficacy of alcohol interventions for first-year college students: A meta-analytic review of randomized controlled trials. Journal of Consulting and Clinical Psychology. 2014; 82(2):177–188.10.1037/a0035192 [PubMed: 24447002]

Sutton AJ, Higgins JPT. Recent developments in meta-analysis. Statistics in Medicine. 2008; 27(5):625–650.10.1002/sim.2934 [PubMed: 17590884]

Terlecki, M. The long-term effect of a brief motivational alcohol intervention for heavy drinking mandated college students. Ph.D., Louisiana State University and Agricultural and Mechanical College; Baton Rouge, LA: 2011.

Walters ST, Bennett ME, Miller JH. Reducing alcohol use in college students: A controlled trial of two brief interventions. Journal of Drug Education. 2000; 30(3):361–372.10.2190/JHML-0JPD-YE7L-14CT [PubMed: 11092154]

Walters ST, Neighbors C. Feedback interventions for college alcohol misuse: what, why and for whom? Addictive Behaviors. 2005; 30(6):1168–1182.10.1016/j.addbeh.2004.12.005 [PubMed: 15925126]

Walters ST, Vader AM, Harris TR, Field CA, Jouriles EN. Dismantling motivational interviewing and feedback for college drinkers: A randomized clinical trial. Journal of Consulting and Clinical Psychology. 2009; 77(1):64–73.10.1037/a0014472 [PubMed: 19170454]

Walton MA, Chermack ST, Shope JT, et al. Effects of a brief intervention for reducing violence and alcohol misuse among adolescents: A randomized controlled trial. Journal of the American Medical Association. 2010; 304(5):527–535.10.1001/jama.2010.1066 [PubMed: 20682932]

White A, Hingson R. The burden of alcohol use: Excessive alcohol consumption and related consequences among college students. Alcohol Research: Current Reviews. 2013; 35(2):201–218. [PubMed: 24881329]

White House. FACT SHEET: President Obama's Precision Medicine Initiative. 2015 Jan. Retrieved from https://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative

White HR, Morgan TJ, Pugh LA, Celinska K, Labouvie EW, Pandina RJ. Evaluating two brief personal substance use interventions for mandated college students. Journal of Studies on Alcohol. 2006; 67:309–317. [PubMed: 16562414]

White HR, Mun EY, Pugh L, Morgan TJ. Long-term effects of brief substance use interventions for mandated college students: sleeper effects of an in-person personal feedback intervention. Alcoholism: Clininical & Experimental Research. 2007; 31(8):1380–1391.10.1111/j.1530-0277.2007.00435.x

```
┌─────────────────────────────┐        ┌─────────────────────────────┐
│   Records identified through │        │  Additional records identified│
│     database searching       │        │    through other sources     │
└─────────────────────────────┘        └─────────────────────────────┘
                │                                      │
                ▼                                      ▼
        ┌─────────────────────────────────────────────┐
        │       Records after duplicates removed       │
        └─────────────────────────────────────────────┘
                              │
                              ▼
                ┌─────────────────────┐        ┌─────────────────────┐
                │   Records screened   │───────▶│   Records excluded   │
                └─────────────────────┘        └─────────────────────┘
                              │
                              ▼
                ┌─────────────────────┐        ┌─────────────────────┐
                │ Full-text articles   │───────▶│ Full-text articles   │
                │ assessed for         │        │ excluded, with reasons│
                │ eligibility          │        └─────────────────────┘
                └─────────────────────┘
                              │
                              ▼
                ┌─────────────────────┐
                │  Studies included in │
                │  qualitative synthesis│
                └─────────────────────┘
                              │
                              ▼
                ┌─────────────────────┐
                │  Studies included in │
                │ quantitative synthesis│
                │   (meta-analysis)    │
                └─────────────────────┘
```
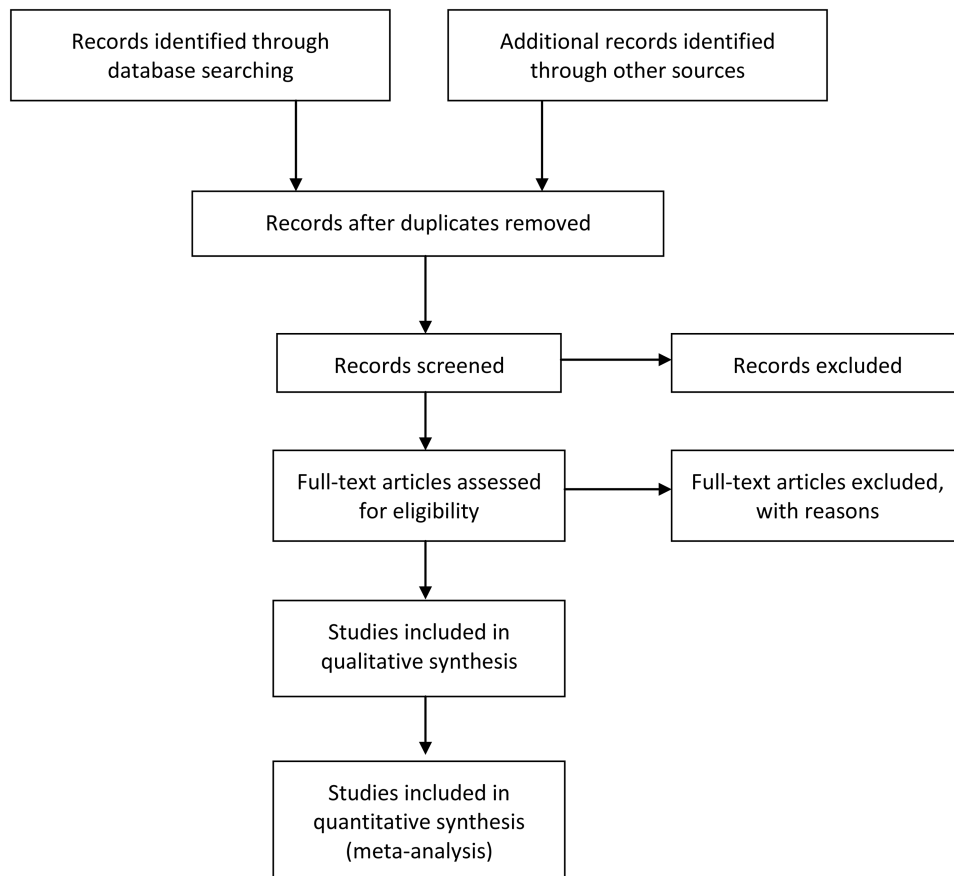
Figure 1.
The PRISMA (2009) flow chart across stages of a systematic review: Study identification, screening, eligibility, and inclusion.
Modified from Moher et al. (2009). In the original PRISMA diagram, sample size for each rectangular box is included.
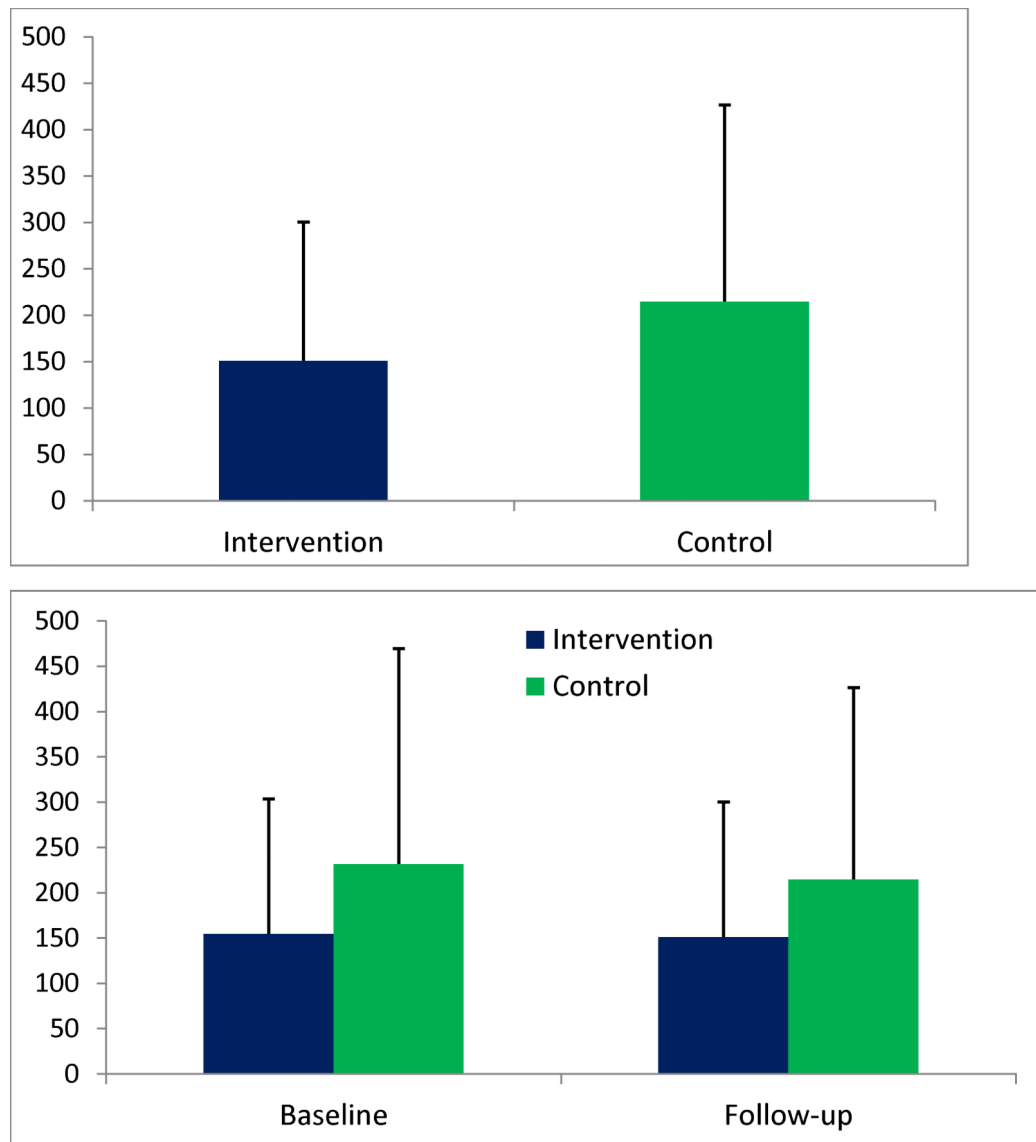
Figure 2.

a. Motivational Interviewing intervention had a statistically significant effect on the amount of alcohol consumed in gram (*Y*-axis) during a typical weekend (Friday night-Sunday night) in Marsden et al. (2006), according to Foxcroft et al.'s data (*SMD* = −0.35; *SE* = 0.11 at 6-month follow-up). Error bars indicate standard deviations.

b. In truth, no evidence of any change or intervention effect existed in Marsden et al. (2006). Error bars indicate standard deviations.

Table 1

Summary of the Limitations of the Meta-analysis Review (2014) by Foxcroft et al.

Research questions and objectives

- Obtaining overall MI effect sizes was the primary aim. Heterogeneous samples, settings, interventions, designs, and follow-ups were included and combined, without examining between-study heterogeneity in effect sizes. Potentially meaningful subgroups for MI were not explored in the presence of substantial statistical heterogeneity beyond chance.

Study identification, screening, and eligibility evaluation procedures

- Search strategies were not comprehensive.

- The qualifications of reviewers (expertise and training background) were undisclosed.

- Screening was based on titles and abstracts but not full-text articles.

- Multiple studies reporting the same trial data were treated as if they were independent studies.

- Disagreements between reviewers and coders were undocumented.

Data abstraction and coding procedures

- Data from eligible follow-ups from eligible studies were omitted.

- For studies with multiple follow-ups at 4 months and more, outcome data from the final follow-up were chosen over short-term follow-up.

- Inaccurate data extraction and entry:

  – Sample sizes were coded inconsistently and inaccurately;

  – No bias correction was done for small samples;

  – Data for intervention and control groups were switched;

  – Mean and standard deviation estimates were inaccurately coded; and

  – Follow-up time periods were coded incorrectly and data combined incorrectly.

Data analysis, data presentation, and interpretation

- As a result of the flaws noted from the previous stages, effect size estimates and standard errors may not be accurate.

- Data extracted from original studies may be correct but still provide inaccurate data (see Marsden et al. [2006] in Figure 2).

- Complex data structures – multiple related interventions, subgroups, and outcomes – were unaccounted, raising concerns of nonindependence among estimates and inflated sample size.

Table 2

Recalculated or Correct *SMDs* and *SEs* for the 6-month Follow-up Data in Walters et al. (2009)

	N	Mean	SD	SMD	SE	g	SE
Drinks per week							
MIF	67	10.19	8.71				
FBO	54	12.07	12.31	-0.18	0.18	-0.18	0.18
MIO	59	11.59	9.55				
AO	61	12.92	14.16	-0.11	0.18	-0.11	0.18
Alcohol-related problems							
MIF	66	4.06	4.96				
FBO	54	3.72	4.70	0.07	0.18	0.07	0.18
MIO	59	5.41	7.28				
AO	61	5.77	6.11	-0.05	0.18	-0.05	0.18

Note. MIF = Motivational interviewing with feedback; FBO = Feedback only; MIO = motivational interviewing only; AO = Assessment only. *SMD* = Standardized mean difference; *SE* = Standard error; *g* = Hedge's *g*. The underlined numbers indicate errors in Foxcroft et al. (2014) that reported *SMD (SE)* = 0.07 (0.18) and -0.05 (0.18) for drinks-per-week outcome data. *N = 66* for alcohol-related problems for MIF was incorrectly reported as 67 in Foxcroft et al. Means and standard deviations were not reported in Foxcroft et al. but were obtained directly from Table 1 in Walters et al. (2009).