A Corrective Meta-Analysis of Personalized Normative Feedback

David Zimmerman[a,*], Baruch Fischhoff[a,b]

[a]Department of Social and Decision Sciences, Carnegie Mellon University
[b]Department of Engineering and Public Policy, Carnegie Mellon University

Abstract

Personalized Normative Feedback (PNF) has been proposed as an inexpensive, scalable intervention for reducing problematic consumption of alcohol, particularly among college students. Many individual studies, as well as meta-analyses, have tested the efficacy of PNF. The findings have been generally positive, demonstrating that it decreases alcohol consumption and the problems associated with excessive consumption. Unfortunately, many of these studies have less than ideal methodologies, which potentially introduce bias to their results. We apply a quantitative adjustment procedure to the findings of each study to account for these biases. Results were divided by a factor of 1.61 on average. While many of the results remain statistically significant after correction, the effects are relatively small, less than 0.2 (Cohen's *d*), in all. Other methods will need to be developed if PNF is to achieve dramatic progress towards reducing drinking. Little evidence exists on the long-term impacts of PNF, or how PNF interacts with the transition out of college.

*Corresponding author. Phone 1-206-790-9002.
Email: zimmerman [dot] david [dot] b [at] gmail.com.

Section 1: Introduction

Problematic alcohol consumption has numerous detrimental effects on society. It reduces the health (White & Jackson, 2004) and productivity (Anderson & Baumberg, 2006; Smit et al., 2006) of individuals. Although lay theories predict that alcoholics create the greatest harmful to society, problem drinkers (e.g., bingers) have a much larger negative impact (Heather & Kaner, 2003; Kaner et al., 2007). Unfortunately, a disproportionate number of college students drink heavily and consume alcohol at dangerous levels (Schulenberg & Maggs, 2002).

Personalized normative feedback (PNF) has been studied intensively as a method for decreasing alcohol consumption, particularly among college students. It involves telling individuals the actual alcohol consumption of a referent group. Being aware that personal consumption or that estimates of average consumption are higher than the actual average is key to PNF working. The procedure is intended to use social norms as the mechanism for change (Berkowitz, 2005). PNF assumes that participants overestimate the alcohol consumption of the reference group. Social norms theory suggests that by making individuals aware that they either consume much more than the average or think the average is much higher the reality, they will be influenced to consume less.

Any salient referent group can be used according to theory, with little work analyzing which reference groups are the most effective. Two common reference groups are all college students in the country or college students at the university where the intervention is taking place. Ideally, normative information would reflect the major determinants of population variation. Gender is foremost of these, because men drink more than women based on numerous measures, such as how much they drink, the frequency of consumption, and drinking to get drunk (Wechsler et al., 2001). With overall norms, many women will discover that they drink less than the actual average consumption of the referent group or perceive all college students as a poor reference group, reducing any pressure from social norms to consume less.

Heavy or binge drinkers are commonly targeted with PNF. There are two common ways to classify drinking as heavy: men who drink 5 or more drinks and women who drink 4 or more drinks in a single occasion, or as men who drink more than 20 drinks or women who drink more than 13 drinks in a week. Studies typically ask about alcohol consumption in the previous 2-6 weeks, thus any instance of heavy drinking, be it per occasion or per week, will qualify a participant as a heavy drinker. Because overestimation in normative alcohol consumption is associated with problematic drinking in college students (Clapp & McDonnell, 2000), it might be possible to reduce problematic consumption by changing the perceived norms. Indeed, perceived campus norms are predictive of consumption, even after controlling for the attitude of college students towards consumption (H. W. Perkins, 2007; H. Perkins & Wechsler, 1996).

Several reviews and meta-analyses have found evidence for the efficacy of feedback as a tool for reducing alcohol consumption (Carey, Scott-Sheldon, Carey, & DeMartini, 2007; Larimer & Cronce, 2007; Lewis & Neighbors, 2006; Moreira, Smith, & Foxcroft, 2010; Riper et al., 2009). Many of the reviewed studies include treatments that combine feedback with other information, such as the personal risk of developing alcohol dependency (Lewis & Neighbors, 2006). Only a few studies have had just a PNF arm in a randomized controlled trial. In one, Mattern (2004) found that, among students exposed to a social marketing campaign, changes in consumption were strongly associated with changes in perceived norms. The students who reported a decrease in perceived drinking norms were much more likely to report decreased consumption. There was even an effect whereby students who perceived the norms to have increased were more likely to report an increase in their own consumption.

Brief motivational interventions for alcohol use were originally developed as a method to combat alcoholism. Between 1960-1970, about 10% of the US population had trouble with alcohol consumption (Cahalan, 1970; Moore & Gerstein, 1981). Many cases of problematic consumption were discovered during general screening procedures (Luckie, White, Miller, Icenogle, & Lasoski, 1992). Once these issues were discovered, patients were usually recommended to meet with a specialist and asked to come back at a later date. Patients were scheduled for an initial counseling session about their alcohol consumption, with additional visits if necessary. Unfortunately, few of these individuals returned for any follow-up visits. Typically 5-6% of scheduled consultations were completed (Chafetz, 1961, 1968; Luckie et al., 1992). One solution that emerged was to offer patients with problematic consumption a meeting with a trained counselor immediately after their general appointment. This dramatically increased the follow-up rate, to over 60% (Chafetz, 1961, 1968). Further studies tested whether a short intervention immediately following a traditional visit to a doctor would be effective at reducing problematic consumption relative to the traditional follow-up visit at a future date. The results suggested that the short visits were at least equivalent, and potentially more beneficial for relevant outcome measures (Bien, Miller, & Tonigan, 1993). These findings suggested that brief interventions were the most cost-effective means for reducing alcohol consumption in problematic cases (Holder, Longabaugh, Miller, & Rubonis, 1991).

Brief motivational interventions were redesigned after discovering that many heavy drinkers are unaware of the true average consumption. In fact, they typically overestimate it (Clapp, 2000; Perkins, 1996). Numerous studies looked into the possibility of staging a brief motivational intervention that depended on changing perceived norms, thus utilizing pressure from social norms to decrease individuals' consumption (Riper et al., 2009). This type of intervention was very cost effective because materials could be sent to all individuals in at-risk populations and without requiring meeting a trained counselor in person.

College students are more at-risk for engaging in problematic alcohol consumption relative to similar aged individuals who are not in school (Substance Abuse and Mental Health Services Administration, 2006). Various social norms interventions have been created for this group. For example, Kypri et al. (2003) examined the medium through which college students prefer learning about normative consumption. A recent Cochrane Review meta-analysis of social norm interventions found mixed evidence for the effectiveness of PNF on self reports of alcohol problems, frequency of consumption, quantity, blood alcohol content (BAC), and drinking norms (Moreira et al., 2010). BAC cannot be reasonably self-reported; therefore, researchers ask participants how many drinks they have had on each drinking occasion and how long they were drinking, from which researchers calculate BAC.

Three meta-analyses (Carey et al., 2007; Moreira et al., 2010; Riper et al., 2009) have already investigated this general domain of interventions to decrease alcohol consumption. Meta-analyses generally convert the results of a study into effect sizes before aggregation.

Psychological studies often use mean difference between groups as the outcome of interest and thus use d-prime (i.e., Cohen's *d*) as the metric of effect size. D-prime is the estimated distance between the control and treatment distributions measured in standard deviations. To calculate it, the difference in means is divided by the standard deviation. If the standard deviations are not significantly different for the treatment and control group then a pooled standard deviation is used. An effect size of 0.1-0.3 is considered small, approximately 0.5 is considered medium, and 0.8 or higher is large (Cohen, 1998). One additional property of d-prime is that it replaces the units of the outcome variable with standard deviations. This allows for the weighted averaging of effect sizes for similar outcomes, such as frequency of consumption, without requiring all studies to use the same scale.

Carey et al. (2007) analyzed individual level interventions, not just PNF. Moreira et al. (2010) focused on social norms interventions, both PNF and social norms campaigns, on college or university students. Riper et al. (2009) looked at personalized feedback. All three used fixed and random effects models for their aggregation, reported which model was most appropriate for the specific outcome variable. For effectiveness, Moreira et al. found the largest effect sizes, ranging from 0.24 for alcohol related problems in the 4-16 month follow-up period to 0.77 for peak BAC in follow-ups less than 4 months. Carey et al.'s estimated effect sizes are generally smaller and have narrower confidence intervals for d-prime. The effect sizes range from 0.11 to 0.41; in this case, small to moderate. Riper et al. provides only one effect size, 0.22, which is within Carey et al.'s range. While Riper et al. focus on personalized normative feedback interventions, both Carey et al. and Moreira et al. look at broader categories of interventions. Carey et al. consider studies with motivational interviews, alcohol/BAC education, normative comparisons, feedback on consumption, moderation strategies, feedback on problems, goal setting, or feedback on expectancies. They find that, in general, the effects of the interventions are not lasting and tend to attenuate to the null after 27 weeks. Moreira et al. found web-based social norm interventions as the most effective, with effect sizes ranging from 0.26 to 0.77. Slightly smaller results were found for individual face-to-face interventions. For both web-based and face-to-face intervention methods, Moreira evaluated fewer studies, leading to wider confidence intervals than those in Carey et al. or Riper et al. The effect sizes for Moreira et al. were strongest for the web-based interventions. Riper et al. is not included in the table below because they only produced one effect size (0.22).

Effect Sizes for Previous Meta-Analyses

Short-Term Follow-Up		
Outcome	Carey (2007)	Moreira (2010)[a]
Average Quantity	0.19	0.29
Frequency Hvy Drink[1]	0.17	0.47
Frequency of Drinking[2]	NS	0.38
Peak BAC	0.41	0.77
Drinks Per Occasion	NS	-
Alc Rel Problems[3]	NS	0.31
Medium-Term Follow-Up		
Average Quantity	0.11	NS
Frequency Hvy Drink[1]	0.11	0.22
Frequency of Drinking[2]	NS	0.31
Peak BAC	NS	NS
Drinks Per Occasion	0.19	-
Alc Rel Problems[3]	0.22	0.26
Long-Term Follow-Up		
Average Quantity	NS	NS
Frequency Hvy Drink[1]	NS	-
Frequency of Drinking[2]	0.16	-
Peak BAC	NS	NS
Drinks Per Occasion	NS	-
Alc Rel Problems[3]	0.14	-

NS: No significant change; "-": not measured in the meta-analysis; [a]: web-based interventions; [1]: Frequency of Heavy Drinking; [2]: Frequency of any level of drinking; [3]: Alcohol Related Problems

Table 1: Cohen's *d* effect sizes from previous meta-analyses of PNF with multiple outcome variables

Larimer and Cronce's (2007) qualitative review findings are congruent with these formal meta-analyses. They look at interventions directed at individuals as opposed to broader marketing campaigns. Several PNF interventions involved posters in dorm halls, articles in the newspaper, and presentations open to all students. PNF was generally successful at reducing alcohol consumption, while other strategies, such as education and awareness or cognitive and behavioral skills programs, produced more mixed results. Larimer and Cronce call for studies with greater methodological rigor, such as true random assignment, larger sample sizes, assessment-only control conditions, in-person normative interventions, and BAC training. Lewis and Neighbors (2007) looked at subgroup effects in PNF studies, finding that the intervention appears most effective on heavy drinkers. Some campus sub-groups (e.g., Greek members, dormitory residents, freshmen, and athletes) are more likely to engage in heavy drinking and have alcohol related problems. However, PNF has not been particularly successful with them. Lewis and Neighbors suggest using different reference groups, for example, friends specific, gender specific, group specific, and age specific groups.

All of these summaries accept the studies involved as methodologically sound. Carey et al. did not discuss methodological strength. Riper et al. coded studies in terms of there methodological features: allocation of condition by a third party, random allocation concealment on respondents, blinding of assessors of outcomes and attrition. However, this assessment of methodological quality did not affect any of the results from the aggregation procedure.

Even if their aggregation procedures are appropriate, if the studies are biased, so will be their estimates of the efficacy of PNF interventions. The Cochrane Review process has identified a set of methodological features as fundamental to producing valid scientific findings, violation of which increases researchers' chance of finding the results they seek, hence increase Type I errors (Jüni, Altman, & Egger, 2001; Kjaergard, 2001; K. Schulz, Chalmers, Hayes, & Altman, 1995; Stukel & Fisher, 2007). Moreira et al. (2010) discuss how some of the studies in their review have many of these biases (Higgins et al., 2011)

Cochrane created a coding scheme to systematically rate the quality of a study in these terms. However, these concerns are not reflected in summary meta-analyses, except by occasionally omitting studies considered to be particularly flawed. No formalized process was implemented to determine which studies to omit. However, several meta-analyses have estimated the bias introduced in clinical trials from particular methodological flaws (Jüni et al., 2001; Kjaergard, 2001; K. Schulz et al., 1995; Stukel & Fisher, 2007). By collecting many studies, some with and some without a bias they could compare the findings and estimate its impact.

Davis it al. (2012) proposes a procedure for systematically correcting for such biases, based on risk-of-bias analysis. Each study is coded for methodological flaws. When they have these flaws, the variance and magnitude of their results are adjusted according to the estimates derived in meta-epidemiological studies. These adjustments typically decrease the strength of the findings and increase the variance of each study, both reducing the significance of their results.

We compiled interventions using PNF to decrease alcohol consumption in college students as well as the general population. With this set of studies, we ran a corrective meta-analysis, adjusting reported results using the procedure from Davis et al. (2012). We find sufficiently high rates of these flaws to reduce estimates of the effects of PNF interventions below the small ones reported in the reviews without corrections. Section 2.1 details the collection and selection process for included studies on PNF. Section 2.2 presents the biases, how we scored the studies for them, and the adjustment process. Section 2.3 explains the specific methods used for aggregation and sensitivity analysis of the findings. Section 3.1 details how our analyses are reported. Section 3.2 lists all of the meta-analyses for follow-up periods from 0-3 months. Section 3.3 lists all meta-analyses with follow-up periods from 4-16 months. Section 4.1 discusses the primary results from the meta-analyses. Section 4.2 discusses the reporting quality of included studies.

Section 2: Methods

Section 2.1 Collection and Selection of Included Studies

Studies were gathered from several sources including Internet databases such as Web of Science, Proquest, Sage Journals Online, Science Direct, and Google Scholar; pertinent references from articles found; and online articles from professional journals. The Internet databases were queried with a combination of "alcohol," "personalized normative feedback," "social comparison," "social norms," and "intervention." Included studies had to use social comparison feedback as part of one or more interventions, be a randomized controlled trial or quasi-experimental procedure, include a control or reasonable placebo condition, measure alcohol consumption, provide the information needed to calculate effect sizes, and be appropriate for random and fixed effects analyses (see Table 2 for details). A study was excluded if it used an active control condition or combined interventions directed at substances other than alcohol. Most of the studies come from the previous reviews and meta-analyses.

Requirements of Included Studies

Category	Specifications
Social Comparison Feedback	(1) Use of referent group for normative feedback
Proper assignment of condition	(1) Random assignment (2) Quasi-experiment
Proper control condition	(1) Assessment only condition (2) Placebo condition
Measure alcohol consumption	
Provide values to calculate effect sizes	(1) Within group means, standard deviations, and sample sizes (2) Relevant statistical tests or confidence intervals
Appropriate for fixed and random effects meta-analyses	(1) Outcomes can be transformed into normal distributions (2) Randomization occurs at individual level

Table 2: Requirements for PNF intervention to be included in current meta-analysis

In total, 56 studies were identified as potential candidates. Thirteen were eliminated for not having an assessment only control or reasonable placebo condition. Another two did not measure alcohol consumption, five were not appropriate for random and fixed effects models, and six did not use normative feedback. Random and fixed effects models assume that the data are analyzed at the level that randomization occurred and the response variables are normally distributed or can be transformed to be so. These five studies did not randomize at the individual level, making random and fixed effects models inappropriate.

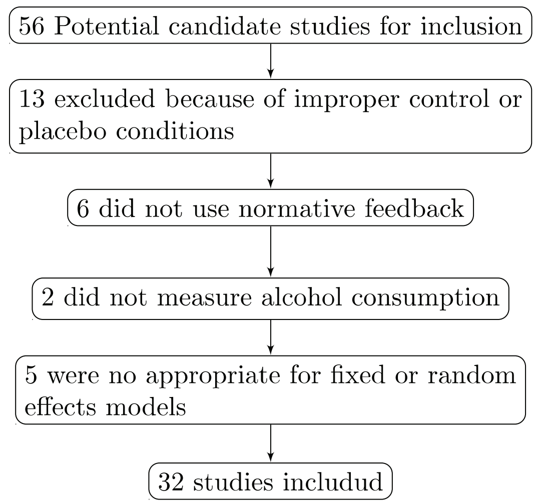


Figure 1: Flow chart of PNF studies to be included in current meta-analysis

Several relevant outcome variables were commonly reported and included for analysis here. Many studies measured some combination of frequency of consumption, general average consumption (e.g., average weekly consumption), peak BAC, alcohol related problems, quantity consumed per occasion levels, and frequency of heavy drinking. All included studies reported within-group variances, means, and treatment group-level sample sizes or these values could be found in another source, such as a meta-analysis or reported confidence intervals. Analyses of effects were broken down by outcome variable to test the specific benefits of PNF. For example, all of the studies measuring heavy drinking were combined in their own meta-analysis. Separating the results by studies conducted on the same population with the same intervention conditions also prevents inappropriate covariance. Different outcomes measured on the same population in the same experimental setup are not independent, and would thereby break a model assumption if aggregated together in the same analysis. Not all studies used the same units for measuring a particular outcome variable, but the random and fixed effects models convert all findings to a common scale, effect sizes, before aggregation.

Section 2.2: Risks-of-Bias

The Cochrane Collaboration has established rigorous tools for assessing the evidence in medical interventions, primarily for conducting reviews and meta-analyses. One major focus of this toolset is grading criteria for methodological characteristics of in intervention. In particular, the Collaboration highlights methodological flaws that create potential risk of biases. The goal is to correct for any artifactual increase in type I errors. Coders used the table below to determine whether a particular study suffered from one of six biases. *Volunteer selection bias* occurs when volunteers self-select to participate which can be problematic because they may be more willing to change their behavior; *intervention selection bias* occurs when participants are allowed to choose the particular treatment arm they will be placed in; *sequence generation bias* arises when conditions are assigned without using a truly random process; *allocation concealment bias* occurs when the assignments are not hidden from the participants or researchers so that either party can manipulate the condition assigned; *blinding bias* happens when researchers know what treatment group a participant is in and can change how the researcher treats the participant or the interpretation of the participant's behavior; *attrition bias* is when a participant's outcome is associated with the likelihood of being lost to follow-up.

Rules for risk-of-bias classification

Risk-of-Bias	High Risk	Low Risk
Volunteer	Opt-in design	(1) Opt-out design; (2) Mandatory participation; (3) Heckman correction[a]
Intervention	(1) Random assignment before volunteering (allowing withdrawal); (2) Participant or researcher choice; (3) Availability of intervention (4) Assignment based on protests or baseline data	(1) Random assignment after volunteering; (2) Propensity score adjustment[b]
Generation	Alternating, day of birth, sequential, other non-random sequence	Truly Random sequence
Concealment	Not central randomization or similar procedure	Central randomization[c]
Blinding	Participants knew about other intervention groups when recruited	Participants were not informed about alternative intervention or control groups
Attrition	Data exclusions or withdrawals, and data not missing at random	(1) No dropouts or exclusions; (2) Intention-to-treat analysis[d]; (3) Appropriate imputation[e]

(a): The Heckman Correction (Heckman, 1979) statistically controls for factors affecting individuals' chance of being in the sample. (b): Propensity adjustment statistically models factors that lead participants to choose an intervention program (Gelman and Hill, 2007; Wooldridge, 2002). (c): Central randomization is done by a third party (Higgens et al., 2011). (d): Intention-to-treat analysis treats participants in terms of their original treatment assignment, regardless of any subsequent exclusion, non-adherence, or withdrawal (Hollis and Campbell, 1999). (e): Imputation estimates the values of missing data (e.g., by the mean of the non-missing data (Ibrahim et al., 2005)).

Table 3: Rules for coding level of risk for each risk-of-bias category
Reproduced from Davis et al. (2010) pg. 10

Risk of bias categories were coded by two independent individuals. Five topically relevant previously coded studies (Moreira et al., 2010) were used to train the coders. All of the studies included in Moreira et al.'s meta-analysis were not coded because they had already undergone the process. Each article not from the Cochrane review was evaluated using the rules in Table 3 to determine whether a study had low risk, high risk or insufficient information about a bias. These rules adapted ones from previous research (Higgins, Altman, & Sterne, 2011; Turner, Spiegelhalter, Smith, & Thompson, 2009). Each category has a list of characteristics that categorize a study at either high or low risk of bias. For example, a study could be considered at high risk of concealment bias if the procedure did not involve central randomization. For three categories, volunteer bias, intervention bias, and attrition, corrective procedures exist (e.g., intention to treat analysis) that allow statistically adjusting for the bias. If the research report provided insufficient information for a category it was treated as if it had high risk of bias, making this a conservative test of effect sizes. Many studies did not describe the randomization process sufficiently to determine generation bias, concealment bias, and blinding, hence were recorded as high risk-of-bias. The inter-rater reliability of the final sample was kappa = 0.37. All disagreements were resolved through discussion. See table above for all rules for determining risk-of-bias classification.

Risk-of-Bias Codings[a]

Study	Volunteer	Generation	Concealment	Blinding	Attrition
Agostinelli (1995)	Low	High	High	Unclear	High
Bendtsen (2012)	Low	Low	Low	Low	High
Bewick (2008)	High	Low	Low	High	High
Bewick (2013)	High	Unclear	Low	High	High
Borsari (2000)[1]	Low	Low	Unclear	Unclear	High
Borsari (2005)[1]	Low	Low	Unclear	Unclear	Unclear
Carey (2006)[1]	Low	Unclear	Unclear	High	Unclear
Collins (2002)[1]	Low	Unclear	Unclear	Unclear	Unclear
Cucciare (2013)	Low	Low	Low	Unclear	Unclear
Hansen (2012)	Low	Unclear	Low	High	Low
Juarez (2006)[1]	Low	Unclear	Unclear	Unclear	High
Kypri (2004)[1]	Low	Low	Low	Unclear	Unclear
Kypri (2005)[1]	Low	Low	Low	Low	High
Kypri (2008)[1]	Low	Low	Low	Low	Unclear
Kypri (2009)	Low	Low	Low	Low	Low
Lewis (2007a)[1]	Low	Unclear	Unclear	Unclear	High
Lewis (2007b)[1]	Low	Unclear	Unclear	Unclear	High
Marlatt (1998)[1]	Low	Low	Unclear	Unclear	High
Martens (2013)	High	Low	Unclear	Unclear	Unclear
McNally (2003)[1]	High	Low	Unclear	Unclear	High
Michael (2006)[1]	High	Unclear	Unclear	Unclear	Low
Moreira (2012)	High	Low	Low	High	Low
Murphy (2001)[1]	Low	Unclear	Unclear	Unclear	Low
Neal (2004)[1]	Low	Unclear	Unclear	Unclear	Low
Neighbors (2004)	High	Unclear	Unclear	Unclear	Low
Neighbors (2006)[1]	Low	Unclear	Unclear	Unclear	High
Terlecki (2010)	High	Unclear	Unclear	Unclear	High
Voogt (2013)	High	Low	Unclear	Unclear	Low
Walters (2000)[1]	Low	Unclear	Unclear	Unclear	High
Walters (2007)[1]	High	Unclear	Unclear	Unclear	Low
Werch (2000)[1]	High	Unclear	Unclear	Unclear	High
White (2008)	Low	Low	Unclear	Unclear	Low

(a) Intervention risk-of-bias was coded as low for all studies and thus not included in the table. (1) Used in Moreira et al. (2010) Cochrane Review.

Table 4: Coding of each risk-of-bias for all included studies

Based on these codes, study results were corrected in order compensate for expected bias (Stukel et al., 2007; Juni et al, 2001; Kjaergard, 2001; Schulz et al., 1995). Each bias has an adjustment factor representing the anticipated degree of over- or under-estimation of effects. In the table below, adjustment factors for four of the six biases are from meta-epidemiology studies (Jüni et al., 2001). These estimates compared studies with sufficient detail to have a low risk of bias with studies having ambiguous or clearly high risk-of-bias. The adjustment factor for attrition was estimated in a meta-analysis of electricity usage from medical studies (Davis, Krishnamurti, Fischhoff, & Bruin, 2012; Kjaergard, 2001; K. F. Schulz, 1995; Stukel & Fisher, 2007). No estimate for the over- or under-estimation of effect sizes due to volunteer risk-of-bias exists. When studies had more than one bias, the adjustment factors were multiplied together, assuming independence. There is preliminary evidence suggesting the biases introduced by each methodological flaw are independent (Turner et al., 2009), but additional research is needed on this question. For example, a study with insufficient concealment and blinding would have the effect size divided by 1.48 = (1.30 *1.14). Variances were also adjusted when there was high risk of bias. Within-group variances were increased in proportion to the total adjustment to the effect size for risk of biases. Studies with high risk-of-bias, therefore, had smaller effect sizes and increased variance.

Estimates of bias for each type of bias

Source	Bias Type	Bias Estimate	95% CI	Variance
-	Volunteer	-	-	-
Stukel et al. (2007)	Intervention	44%	[19, 69]	156%
Jüni et al. (2001)	Generation	19%	[-9, 40]	196%
Jüni et al (2001)	Concealment	30%	[20, 38]	25%
Jüni et al. (2001)	Blinding	14%	[1, 26]	43%
Kjaergard (2001) Schulz et al. (1995)	Attrition	-8%	[-21, 6]	49%

Table 5: Estimates of bias for each category of bias
Reproduced from Davis et al. (2010) pg. 11

In the original papers: Stukel et al. (2007), Juni et al. (2001), Kjaergard (2001), and Schulz et al. (1995), estimating the correction factor, the conversion from their estimated ratios to adjustment factors was slightly incorrect. Each article computed the ratio of effect sizes for studies without the biases over the ones with them. For example, for concealment bias the ratio of odds ratios was 0.7. In estimating the effect of concealment bias studies without biases could have an average odds ratio of 2.0, and the biased studies have an average odds ratio of 2.86 (0.7 = 2.0/2.86). However, when reporting the percentage change in effect size, they said the intervention outcomes were lower (more beneficial) than for the control groups by a particular percentage in addition to the ratio of odds ratios. Percentages were calculated as the estimated ratio minus one. Concealment bias had an estimated ratio of 0.7, which corresponds to a 30% decrease (-0.3 = 0.7 – 1) in outcomes for biased studies relative to those unbiased studies. This is not the same as stating that relative to an unbiased study the effects are 30% greater than they should be. To correctly account for each bias the effect size would have the percentage of the reported effect size subtracted from itself (1 – (1*0.3) = 0.7). Adjustment factors have been recalculated according to the inverse of the ratio of effect sizes found in the meta-analyses (1/0.7 = 1.43 = +43%). All bias estimates were recalculated from Davis et al. (2012) and can be seen in the table below. Due to this discrepancy both the adjustment factors from Davis et al. (2012) and the newly computed values are used in the meta-analysis aggregation procedures. Values from the Davis et al. procedure are considered moderate adjustments and the newly computed figures are considered the full correction.

Adjusted Estimates of bias for each type of bias

Source	Bias Type	Bias Estimate	95% CI	Variance
-	Volunteer	-	-	-
Stukel et al. (2007)	Intervention	79%	[23,223]	156%
Jüni et al. (2001)	Generation	23%	[-8,67]	150%
Jüni et al. (2001)	Concealment	43%	[25,61]	20%
Jüni et al. (2001)	Blinding	16%	[1, 35]	39%
Kjaergard (2001)	Attrition	-7%	[-17, 6]	46%
Schulz et al. (1995)				

Table 6: Estimates of bias accounting for new percentages for each category of bias

Section 2.3: Modeling of Effects

The Generic Inverse Variance (GIV) meta-analysis method was used to combine the studies. In order to test for robustness, both fixed and random effects models were applied. When heterogeneity is high, the random effects model is more appropriate because it allows for random intercepts for each study. Fixed effects modeling assumes that all effect sizes come from the same distribution. Generally the models had a mixture of low and high heterogeneity; therefore, fixed and random effects models were reported when appropriate. If heterogeneity is low, then each study is weighted primarily by the inverse of its variance, meaning that more stable groups contribute more to the overall mean. When heterogeneity is high, each study is weighted more equally. The Q statistic, a measure of heterogeneity, was used to test statistically whether each subgroup of studies was best modeled by fixed or random effects. Forest plots for the GIV results are in Appendix C.

In order to isolate the effects of personalized normative feedback separate analyses were run for 2 follow-up periods for 1-3 months and 4-16 months. Moreira et al. (2010) used this classification scheme for follow-up. They considered 1-3 months short term and 4-16 months medium. If a study had multiple follow-up responses within one of these time periods then the longest interval from intervention to outcome measurement was used.

Section 3: Results

Section 3.1 Analysis Reporting Style

The studies were grouped by outcome variable and follow-up period. All results from follow-up periods between 0 and 3 months are reported first, followed by follow-up periods greater than 3 months. Each analysis was done with three different adjustment states: none, moderate, and full.

Moreira et al. (2010) also broke up the analysis by communication method. This unfortunately created several meta-analyses that aggregated one or only a few studies. As a result, we choose not to divide studies on this variable. Each analysis includes a test of heterogeneity with the Q statistic. It is distributed as a Chi-square with (k–1) degrees of freedom. The null hypothesis is that the study results are homogeneous. Table 7 summarizes the results and compares them against the effect sizes found in Carey et al. (2007) and Moreira et al. (2010).

Section 3.2: Follow-Up Period of 0-3 Months

Frequency of Consumption

Fourteen studies with a total of 4,121 participants measured frequency of alcohol consumption. Random effects modeling is most appropriate (Q (13) = 27.49, p-value = 0.01) for the no-adjustment analysis. There is a significant effect of the treatment (SMD = -0.19, 95% CI [-0.30, -0.08]). Both the moderate and full adjustment analyses were best modeled with fixed effects (Q(13) = 14.31, p-value = 0.35), and (Q(13) = 12.01, p-value = 0.53) respectively. Each found a significant effect of the intervention on frequency of consumption: moderate adjustment (SMD = -0.14, 95% CI [-0.20, -0.08]) and full adjustment (SMD = -0.14, 95% CI [-0.20, -0.07]).

Average Consumption

Twenty-three studies measured average consumption with a total of 6,803 participants. Without any bias adjustment, the random effects model is most appropriate (Q(22) = 47.2, p-value 0.0014). This analysis found a significant effect of the interventions on average consumption (SMD = -0.19, 95% CI [-0.27, -0.10]). When either adjustment was applied, the fixed effects model fits better (Moderate: (Q(22) = 25.42, p-value = 0.28), Full: (Q(22) = 22.23, p-value =

0.45). Any adjustment reduced the effect sizes, but they remain significant (Moderate: (SMD = -0.14, 95% CI [-0.19, -0.09]), and Full: (SMD = -0.13, 95% CI [-0.18, -0.08]).

<u>Peak BAC level</u>

Eleven studies with 2,350 participants measured peak BAC levels. All levels of adjusted values are appropriate for a random effects model (None: (Q(10) = 39.86, p-value = <0.0001,), Moderate: (Q(10) = 31.22, p-value = 0.0005), Full: (Q(10) = 29.37, p-value = 0.001). All three analyses showed significant results, with the adjusted analyses reporting effect sizes smaller than the unadjusted estimates (None: (SMD = -0.24, 95% CI [-0.43, -0.05]), Moderate: (SMD = -0.19, 95% CI [-0.36, -0.02]), and Full: (SMD = -0.18, 95% CI [-0.34, -0.012]).

<u>Alcohol Related Problems</u>

Nineteen studies with a total of 4,726 participants measured alcohol related problems. Fixed effects models are appropriate for all three adjustment levels (None: Q(18) =26.72, p-value = 0.08), Moderate: Q(18) = 16.19, p-value = 0.58), and Full: Q(18) = 14.2, p-value =0.72)). None of the analyses found a significant effect of the intervention on reducing alcohol related problems (None: (SMD=-0.005, 95% CI [-0.062, 0.052]), Moderate: (SMD = -0.005 95% CI [-0.062, 0.05]), and Full: (SMD = -0.005, 95% CI [-0.062, 0.05])).

<u>Average Consumption per Occasion</u>

Four studies with a total of 2,161 participants measured average consumption per occasion. Fixed effects models are appropriate for all three adjustment levels (None: (Q(3) = 6.19, p-value = 0.10), Moderate: (Q(3) = 3.81, p-value = 0.28), Full: (Q(3) = 3.45, p-value = 0.33).

Additionally, all three showed significant effects of the intervention on average consumption per occasion (None: (SMD = -0.11, 95% CI [-0.20, -0.027]), Moderate: (SMD = -0.11, 95% CI [-0.19, -0.02], and Full: (SMD = -0.10, 95% CI [-0.19, -0.02])).

<u>Frequency of Heavy Drinking</u>

Thirteen studies with a total of 2,524 participants measured frequency of heavy drinking. All three levels of adjustment are modeled best with fixed effects models (None: (Q(12) = 19.67, p-value = 0.07), Moderate: (Q(12) = 9.97, p-value = 0.62), Full: (Q(12) = 8.22, p-value = 0.76). The unadjusted values find the only significant results of the intervention on frequency of heavy drinking (None: (SMD =-0.10, 95% CI [-0.18, -0.02]), Moderate: (SMD =-0.07, 95% CI [-0.15, 0.09]), and Full: (SMD =-0.06, 95% CI [-0.14, 0.02]). Both the moderate and the fully adjusted analyses report much smaller effect sizes.

Effect Sizes (N) for Meta-Analyses

Short-Term Follow-Up			
Outcome	Carey (2007)	Moreira (2010)[a]	Current Study[b]
Frequency of Drinking[1]	NS (5)	0.38 (2)	0.14 (14)
Average Quantity	0.19 (18)	0.29 (5)	0.13 (23)
Peak BAC	0.41 (5)	0.77 (2)	0.18 (11)
Alc Rel Problems[2]	NS (9)	0.31 (3)	NS (19)
Drinks Per Occasion	NS (19)	-	0.10 (4)
Frequency Hvy Drink[3]	0.17 (5)	0.47 (1)	NS (13)
Medium-Term Follow-Up			
Frequency of Drinking[1]	NS (5)	0.31 (3)	0.14 (11)
Average Quantity	0.11 (19)	NS(4)	0.10 (14)
Peak BAC	NS (12)	NS (1)	0.13 (6)
Alc Rel Problems[2]	0.22 (12)	0.26 (3)	NS (13)
Drinks Per Occasion	0.19 (8)	-	NS (3)
Frequency Hvy Drink[3]	0.11 (12)	0.22 (2)	NS (8)

NS: Not significant finding; "-": not measured in the meta-analysis; [a]: web-based interventions; [b]: The fully adjusted values are reported; [1]: Frequency of any level of drinking; [2]: Alcohol Related Problems; [3]: Frequency of Heavy Drinking

Table 7: Comparison of Fully Adjusted Meta-Analysis Results to Previous Findings

Section 3.3: Follow-Up Period of greater than 3 Months

Frequency of Consumption

Eleven studies with a total of 4,364 participants measured frequency of alcohol consumption. All levels of adjustment are best fit by fixed effects models (None: (Q(10) = 8.64, p-value = 0.57), Moderate: (Q(10) = 3.94, p-value = 0.95), and Full: (Q(10) = 3.17, p-value = 0.98). All three models show a significant effect of the intervention on frequency of alcohol consumption (None: (SMD = -0.17, 95% CI [-0.23, -0.11]), Moderate: (SMD =-0.14, 95% CI [-0.20, -0.08]), and Full: (SMD = -0.14, 95% CI [-0.20, -0.08]). As expected the effect sizes are much smaller for the two adjusted analyses.

Average Consumption

Fourteen studies with a total of 4,737 participants measured average consumption. All three levels of adjustment are fit best by fixed effects models (None: (Q(13) = 12.81, p-value = 0.46), Moderate: (Q(13) = 6.96, p-value = 0.90), and Full: (Q(13) = 5.98, p-value = 0.95)). No matter how the values are adjusted, there is a significant effect of the intervention on average consumption (None: (SMD =-.13, 95% CI [--0.19, -0.07]), Moderate: (SMD = -0.11, 95% CI [-0.16, -0.05]), and Full: (SMD =-0.10, 95% CI [-0.16, -0.04])).

Peak BAC level

Six studies with a total of 1,356 participants measured peak BAC levels. Fixed effects models are appropriate for all three levels of adjustment (None: (Q(5) = 1.97 , p-value = 0.85), Moderate: (Q(5) = 1.29, p-value = 0.94), Full: (Q(5) = 1.29 , p-value = 0.95). All three levels of adjustment produce significant relationships between the intervention and peak BAC levels

(None: (SMD = -0.20, 95% CI [-0.30, -0.09]), Moderate: (SMD = -0.14, 95% CI [-0.25, -0.03]), and Full: (SMD = -0.13, 95% CI [-0.24, -0.02]).

Alcohol Related Problems

Thirteen studies with a total of 3,942 participants measured alcohol related problems. At every level of adjustment, fixed effects models are appropriate (None: (Q(12) = 15.98, p-value = 0.19)), Moderate: (Q(12) = 11.36, p-value = 0.50)), and Full: (Q(12) = 10.14, p-value = 0.60)). Only the first two levels of adjustment show a significant effect of the intervention on alcohol related problems (None: (SMD =-0.08, 95% CI [-0.14, -0.02]), Moderate: (SMD =-0.06, 95% CI [-0.12, -0.001]), and Full: (SMD =-0.06, 95% CI [-0.12, 0.004])).

Average Consumption per Occasion

Three studies with a total of 2,438 participants measured average alcohol consumption per occasion. All three levels of adjusted values are best fit by fixed effects models (None: (Q(2) = 0.06, p-value = 0.97), Moderate: (Q(2) = 0.07, p-value =0.96), and Full: (Q(2) = 0.08, p-value = 0.96)). Additionally, none of the three levels of adjustment show a significant effect of intervention on average consumption per occasion (None: (SMD =-0.067, 95% CI [-0.15, 0.012]), Moderate: (SMD =-0.065, 95% CI [-0.14, 0.015]), and Full: (SMD =-0.06, 95% CI [-0.14, 0.015]).

Frequency of Heavy Drinking

Eight studies with a total of 1,677 participants measured the frequency of heavy drinking. All levels of adjustment are best fit by a fixed effects model (None: (Q(7) = 3.03, p-value =0.88), Moderate: (Q(7) = 3.67, p-value = 0.82), and Full: (Q(7) = 3.67, p-value = 0.82). None of three levels of adjustment show a significant effect of the intervention on frequency of heavy drinking (None: (SMD =-0.09, 95% CI [-0.19, 0.007]), Moderate: (SMD = -0.08, 95% CI [-0.18, 0.015]), Full: (SMD =-0.08, 95% CI [-0.17, 0.019]).

Section 4: Discussion

Section 4.1 Summary of Primary Findings

We conducted a meta-analysis of the effects of personalized normative feedback (PNF) on self-reported drinking behavior. We identified 32 suitable studies, 19 included in a Cochrane review and the rest from other sources. All studies used PNF for at least part of the treatment condition, although most combined PNF with additional information (e.g., harmful effects of over-consumption of alcohol). Over half had an assessment-only condition comparing a PNF group with one receiving no treatment, in terms of changes in drinking behavior. For a few others, the control condition was treatment as usual, which typically involved information on how alcohol affects the body. We did not, however, distinguish among specific communication methods (e.g., mail, web-based) because there were too few studies in each subgroup to produce reliable estimates. Each study was evaluated in terms of its risk of six biases found to affect results of clinical trials in medicine (Higgins et al., 2011; Jüni et al., 2001; Kjaergard, 2001; K. Schulz et al., 1995; Stukel & Fisher, 2007). Where risk-of-bias was found, estimates of effect sizes and variances are adjusted procedures from that literature (Davis et al., 2012). The opportunity to make these adjustments allows using studies that had been excluded from previous meta-analyses based on potential flaws, without giving them undue weight. It also corrects for residual biases in generally sound studies, which were included in previous analyses.

The analyses showed that, as expected, effects were stronger in shorter follow-up periods (0-3 mo) than in longer ones (4-16 mo). For the shorter period, frequency of consumption, average quantity consumed, peak BAC, and average consumption per occasion all decreased significantly, relative to controls. For the longer follow-up period reductions were observed in all of these measures except average consumption per occasion. Overall the findings reveal quite small effect sizes, with none greater than 0.2, and the largest being peak BAC for short-term follow-up (0.18).

On average, studies had an adjustment factor 1.61 for full adjustment and 1.43 for moderate adjustment, (Appendix B provides details). The moderate adjustment factors are all closer to 1 than the full adjustment. Davis et al. (2012) used the moderate adjustment factors in a meta-analysis of electricity pilot studies. We corrected the adjustment factors and came up with slightly more conservative values. Not all studies report the same number of outcome variables. By multiplying the adjustment factor by the number of outcome variables reported we get a weighted average of adjustment factors. This yields an average adjustment factor of 1.60 for the full adjustment and 1.42 for the moderate adjustment. As a result, meta-analyses with moderate and full adjustments produce similar estimates for standardized mean differences. Overall, the adjustment procedure significantly reduced estimated effect sizes. Without the adjustment, both frequency of heaving drinking in the short follow-up period and alcohol related problems in the longer follow-up period show significant decreases relative to controls. With the adjustment neither is. Only peak BAC in the short follow-up period (0-3 mo) has an effect size over 0.2 (0.24) when no adjustments are applied. For the moderate adjustment, none of the effect sizes were over 0.2.

Adjusting for risks-of-bias also had an impact on the heterogeneity of the meta-analyses. In the shorter follow-up period both frequency of consumption and average weekly consumption were best fit by random effects models without any adjustment, but fixed effects models were a better fit with any adjustment. Even in the outcome variables where fixed effects were appropriate for all the models almost all the Q statistic decreased as the adjustment level increased. There is only one case in which the Q statistic went up with adjustment. For frequency of heavy drinking in the 4-16 month follow-up window the Q statistic for no adjustment was 3.03 (df = 7), 3.67 (df = 7) for moderate adjustment, and 3.67 (df = 7) for full adjustment.

Although statistically significant, the effect sizes might still have little clinical relevance. The largest effect size, 0.18, was for peak BAC in the shorter follow-up period. An effect size of 0.18 corresponds to 0.02 units on the BAC scale. If PNF could decrease the average peak BAC by 0.02 there would be few noticeable differences the negative impacts from excessive drinking. For example, decreasing peak BAC would not reduce impairment dramatically (*Alcohol Overdose: The Dangers of Drinking Too Much*, 2013). Approximately 1,825 college students age 18-24 die from alcohol-related injuries, excluding motor-vehicle incidents (Hingson, Zha, & Weitzman, 2009). Individuals are at serious risk of death when they have a BAC level of 0.31-0.45. Reducing this by 0.02 would not produce a major reduction in deaths from alcohol related injuries.

Even though the potential benefits of PNF interventions are quite low, the costs of their basic implementations are low as well. Conducting a PNF intervention requires collecting average consumption data, designing a simple webpage, and emailing its URL to students, checking that they review it. Estimates of national alcohol consumption are readily available, as is the ability to mail links to students already. Getting students to pay attention might be more challenging. Thus PNF interventions have low cost. Several important caveats accompany this basic scheme. If a college wants to collect data from its students and use their drinking behavior as the referent group, then the costs will be higher. Surveys asking about alcohol consumption are complicated. Many people answering the questions will be admitting that they engaged in illegal behavior and may have difficulty remembering their own consumption, particularly if they were highly intoxicated. When asking about sensitive behavior, there is always the concern that people will underreport their activity because of social conformity. Thus, while the cost of implementing a survey on campus may not be prohibitive, the logistical and methodological problems of designing and implementing it might be. Analogous problems would face any larger-scale intervention with a general population. Once a mechanism is in place for distributing PNF individuals must also choose to undergo treatment. Response rates to standard email surveys are around 20% (Kaplowitz, Hadlock, & Levine, 2004). This means that it would be difficult expose the majority of college students to PNF without adding more expensive methods of reaching them. Between how little PNF changes behavior and potentially how prohibitive ideal implementation is, it seems inefficient to implement PNF ubiquitously.

As seen in our analyses, as the time from exposure to these social norms increases, the effects weaken. For example, peak BAC is decreased by 0.18 standard deviations for the short follow-up period, but it decreases to 0.14 standard deviations for the longer period. One method to maintain effectiveness in the long-term would be a reminder. Reminders would continue exposure to normative information and potentially maintain the reductions in drinking behavior.

Given that the decreases in alcohol seem temporary there is another potential hurdle that PNF must overcome, graduation. None of the included studies discussed how interventions might work in the transition period after college students graduate. It is possible that simple exiting the college environment is sufficient to decrease alcohol consumption.

Section 4.2: Reporting Quality of Compiled Studies

The overall quality of reporting, in both the methods and results, of the studied papers appears good. There are a few areas that require additional details, such as how the randomization process for assigning treatments occurred. Of all the risk-of-bias categories coded, 36% were unclear in the published articles. The most common failings were failing to report sufficient details about potential generation, concealment, and blinding biases. In most cases, studies only needed to add a few sentences addressing these concerns and be rescored as low risk-of-bias (if that proved to be the case). As for volunteer bias, all studies described their recruitment process sufficiently to make a determination about the level of risk-of-bias. Unfortunately, there are no accepted corrections for that bias, where it is found. The most common bias was lack of blinding. Twenty-three studies were unclear about the blinding status of their participants, typically arising from insufficient reporting on what participants were told when they began the trials. There are many instances where it is difficult to implement a study with perfect methods because of various logistical factors. The Cochrane Collaboration deliberately chose the term risk-of-bias, in part because many researchers are forced to use less than ideal methodologies. Calling the use of poor methodology flaws can be seen as too harsh a judgment when researchers must use volunteers or are unable to reasonably centrally randomize treatments.

Acknowledgements

I would like to thank my advisor Dr. Baruch Fischhoff for his support and patience while I attempted a project in an unfamiliar domain with a novel methodology. Thanks go to Dr. Stephen Broomell for bolstering my comfort in tackling such an unusual methodology. Thanks go to Amritha Mallikarjun for slogging through the paper to code them for risks-of-bias. Finally, a particular note of gratitude to Tamar Feigenbaum for being the best study buddy one could ask for.

References

Studies Included in the Meta-Analysis

Agostinelli, G., Brown, J., & Miller, W. (1995). Effects of Normative Feedback on consumption among heavy drinking college students. *Journal of Drug Addiction*, *25*(1), 31–40.

Bendtsen, P., McCambridge, J., Bendtsen, M., Karlsson, N., & Nilsen, P. (2012). Effectiveness of a proactive mail-based alcohol Internet intervention for university students: dismantling the assessment and feedback components in a randomized controlled trial. *Journal of Medical Internet Research*, *14*(5), e142. doi:10.2196/jmir.2062

Bewick, B. M., Trusler, K., Mulhern, B., Barkham, M., & Hill, A. J. (2008). The feasibility and effectiveness of a web-based personalised feedback and social norms alcohol intervention in UK university students: a randomised control trial. *Addictive Behaviors*, *33*(9), 1192–8. doi:10.1016/j.addbeh.2008.05.002

Bewick, B. M., West, R. M., Barkham, M., Mulhern, B., Marlow, R., Traviss, G., & Hill, A. J. (2013). The effectiveness of a Web-based personalized feedback and social norms alcohol intervention on United Kingdom university students: randomized controlled trial. *Journal of Medical Internet Research*, *15*(7), e137. doi:10.2196/jmir.2581

Borsari, B., & Carey, K. B. (2000). Effects of a Brief Motivational Intervention With College Student Drinkers. *Journal of Consulting and Clinical Psychology*, *68*(4), 728–733.

Borsari, B., & Carey, K. B. (2005). Two brief alcohol interventions for mandated college students. *Psychology of Addictive Behaviors*, *19*(3), 296–302. doi:10.1037/0893-164X.19.3.296

Carey, K. B., Carey, M. P., Maisto, S. A., & Henson, J. M. (2006). Brief Motivational Interventions for Heavy College Drinkers: A Randomized Controlled Trial. *Journal of Consulting and Clinical Psychology*, *74*(5), 943–954. Retrieved from http://psycnet.apa.org/journals/ccp/74/5/943/

Collins, S., Carey, K., & Sliwinski, M. (2002). Mailed Personalized Normative Feedback as a Brief Intervention for At-Risk College Drinkers. *Journal of Studies on Alcohol and Drugs*, *63*(5), 559–567.

Cucciare, M., Weingardt, K., Ghaus, S., Matthew, B., & Susan, F. (2013). A Randomized Controlled Trial of a Web-Delivered Brief Alcohol Intervention in Veterans Affairs Primary Care. *Journal of Studies on Alcohol and Drugs*, *74*(3), 428–436.

Hansen, A. B. G., Becker, U., Nielsen, A. S., Grønbæk, M., Tolstrup, J. S., & Thygesen, L. C. (2012). Internet-based brief personalized feedback intervention in a non-treatment-seeking population of adult heavy drinkers: a randomized controlled trial. *Journal of Medical Internet Research*, *14*(4), e98. doi:10.2196/jmir.1883

Juarez, P., Walters, S. T., Daugherty, M., & Radi, C. (2006). A Randomized Trial of Motivational Interviewing and Feedback with Heavy Drinking College Students. *Journal of Drug Education*, *36*(3), 233–246.

Kypri, K., Hallett, J., & Howat, P. (2009). Randomized Controlled Trial of Proactive Web-Based Alcohol Screening and Brief Intervention for University Students. *Archives of Internal Medicine*, *169*(16), 1508–1514. Retrieved from http://archinte.ama-assn.org/cgi/reprint/169/16/1508.pdf

Kypri, K., Langley, J., Saunders, J., Cashell-Smith, M., & Herbison, P. (2008). Randomized Controlled Trial of Web-Based Alcohol Screening and Brief Intervention in Primary Care. *Archives of Internal Medicine*, *168*(5), 530–536.

Kypri, K., & McAnally, H. M. (2005). Randomized controlled trial of a web-based primary care intervention for multiple health risk behaviors. *Preventive Medicine*, *41*(3-4), 761–6. doi:10.1016/j.ypmed.2005.07.010

Kypri, K., Saunders, J. B., Williams, S. M., McGee, R. O., Langley, J. D., Cashell-Smith, M. L., & Gallagher, S. J. (2004). Web-based screening and brief intervention for hazardous drinking: a double-blind randomized controlled trial. *Addiction (Abingdon, England)*, *99*(11), 1410–7. doi:10.1111/j.1360-0443.2004.00847.x

Lewis, M., & Neighbors, C. (2007). Optimizing Personalized Normative Feedback: The Use of Gender-specific Referents. *Journal of Studies on Alcohol and Drugs*, *68*(2), 228–237.

Lewis, M., Neighbors, C., Laura, O.-A., Kirkeby, B., & Larimer, M. (2007). Indicated Prevention for Incoming Freshmen: Personalozed Normative Feedbak and High-Risk

Drinking. *Addictive Behaviors*, *32*(11), 2495–2508. doi:10.1016/j.addbeh.2007.06.019.Indicated

Marlatt, G. A., Baer, J. S., Kivlahan, D. R., Dimeff, L. A., Larimer, M. E., Quigley, L. A., … Williams, E. (1998). Screening and Brief Intervention for High-Risk College Student Drinkers: Results From a 2-Year Follow-Up Assessment. *Journal of Consulting and Clinical Psychology*, *66*(4), 604–615.

Martens, M. P., Smith, A. E., & Murphy, J. G. (2013). The efficacy of single-component brief motivational interventions among at-risk college drinkers. *Journal of Consulting and Clinical Psychology*, *81*(4), 691–701. doi:10.1037/a0032235

McNally, A. M., & Palfai, T. P. (2003). Brief group alcohol interventions with college students: Examining motivational components. *Journal of Drug Education*, *33*(2), 159–176. Retrieved from http://baywood.metapress.com/index/82CTLRC5AMTWC090.pdf

Michael, K. D., Curtin, L., Kirkley, D. E., Jones, D. L., & Harris, R. (2006). Group-based motivational interviewing for alcohol use among college students: An exploratory study. *Professional Psychology: Research and Practice*, *37*(6), 629–634. doi:10.1037/0735-7028.37.6.629

Moreira, M. T., Oskrochi, R., & Foxcroft, D. R. (2012). Personalised normative feedback for preventing alcohol misuse in university students: Solomon three-group randomised controlled trial. *PloS One*, *7*(9), 1–10. doi:10.1371/journal.pone.0044120

Murphy, J. G., Duchnick, J. J., Vuchinich, R. E., Davison, J. W., Karg, R. S., Olson, A. M., … John, W. (2001). Relative Efficacy of a Brief Motivational Intervention for College Student Drinkers. *Psycholody of Addictive Behaviors*, *15*(4), 373–379. doi:10.1037//0893-I64X.15.4.373

Neal, D. J., & Carey, K. B. (2004). Developing discrepancy within self-regulation theory: Use of personalized normative feedback and personal strivings with heavy-drinking college students. *Addictive Behaviors*, *29*(2), 281–297. doi:10.1016/j.addbeh.2003.08.004

Neighbors, C., Larimer, M. E., & Lewis, M. a. (2004). Targeting misperceptions of descriptive drinking norms: efficacy of a computer-delivered personalized normative feedback intervention. *Journal of Consulting and Clinical Psychology*, *72*(3), 434–47. doi:10.1037/0022-006X.72.3.434

Neighbors, C., Lewis, M. A., Bergstrom, R., & Larimer, M. E. (2006). Being Controlled by Normative Influences: Self-Determination as a Moderator of a Normative Feedback Alcohol Intervention. *Health Psychology*, *25*(5), 571–579.

Terlecki, M. a, Buckner, J. D., Larimer, M. E., & Copeland, A. L. (2012). Brief motivational intervention for college drinking: the synergistic impact of social anxiety and perceived drinking norms. *Psychology of Addictive Behaviors*, *26*(4), 917–23. doi:10.1037/a0027982

Voogt, C., Poelen, E. A. P., Kleinjan, M., Lemmers, L. A. C. J., & Engels, R. C. M. E. (2013). The Effectiveness of the “what do you drink” web-based brief alcohol intervention in reducing heavy drinking among students: a two-arm parallel group randomized controlled trial. *Alcohol and Alcoholism*, *48*(3), 312–321.

Walters, S., Bennett, M., & Miller, J. (2000). Reducing alcohol use in college students: a controlled trial of two brief interventions. *Journal of Drug Education*, *30*(3), 361–372. Retrieved from http://baywood.metapress.com/index/JHML0JPDYE7L14CT.pdf

Walters, S. T., Vader, A. M., & Harris, T. R. (2007). A controlled trial of web-based feedback for heavy drinking college students. *Prevention Science*, *8*(1), 83–8. doi:10.1007/s11121-006-0059-9

Werch, C., & Pappas, D. (2000). Results of a social norm intervention to prevent binge drinking among first year college students. *Journal of American College Health*, *49*(2), 85–92. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/07448480009596288

White, H. R., Mun, E. Y., & Morgan, T. J. (2008). Do brief personalized feedback interventions work for mandated students or is it just getting caught that works? *Psychology of Addictive Behaviors*, *22*(1), 107–16. doi:10.1037/0893-164X.22.1.107

Additional References

Alcohol Overdose: The Dangers of Drinking Too Much. (2013).

Anderson, P., & Baumberg, B. (2006). *Alcohol in Europe. A public health perspective*. London.

Berkowitz, A. (2005). An overview of the social norms approach. In Cresskill (Ed.), *Changing the Culture of College Drinking: A Socially Situated Health Communication Campaign*. NJ: Hampton Press.

Bien, T. H., Miller, W. R., & Tonigan, J. S. (1993). Brief interventions for alcohol problems: a review. *Addiction (Abingdon, England)*, *88*(3), 315–35. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8461850

Cahalan, D. (1970). *Problem Drinkers: A National Survey*. San Francisco: Jossey-Bass.

Carey, K. B., Scott-Sheldon, L. a J., Carey, M. P., & DeMartini, K. S. (2007). Individual-level interventions to reduce college student drinking: a meta-analytic review. *Addictive Behaviors*, *32*(11), 2469–94. doi:10.1016/j.addbeh.2007.05.004

Chafetz, M. E. (1961). A procedure for establishing therapeutic contact with the alcoholic. *Quarterly Journal of Studies on Alcohol*, 325–328.

Chafetz, M. E. (1968). Research in the alcohol clinic: an around-the-clock psychiatric service of the Massachusetts General Hospital. *American Journal of Psychiatry*, *124*, 1674–1679.

Clapp, J. D., & McDonnell, A. L. (2000). The relationship of perceptions of alcohol promotion and peer drinking norms to alcohol problems reported by college students. *Journal of College Student Development*, *41*(1), 19–26.

Cohen, J. (1998). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Taylor & Francis.

Davis, A., Krishnamurti, T., Fischhoff, B., & Bruin, W. de. (2012). Setting a standard for electricity pilot studies. Retrieved from http://openwetware.org/images/4/45/Davis,_A._Setting_a_standard.pdf

Gelman, A. (2007). *Data analysis using regression and multilevel/hierarchical models*. Retrieved from http://books.google.com/books?hl=en&lr=&id=c9xLKzZWoZ4C&oi=fnd&pg=PR17&dq=Data+Analysis+Using+Regression+and+Multilevel+/+Hierarchical+Models&ots=b9U3Q-Sqme&sig=xe5htWqUmhuqiDMq6lDJ2MT-2Fo

Heather, N., & Kaner, E. (2003). Brief interventions against excessive alcohol consumption. In D. Warell, T. Cox, J. Firth, & E. Al (Eds.), *Oxford textbook of medicine*. Oxford: Oxford University Press.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161.

Higgins, J., Altman, D., & Sterne, J. (2011). Chapter 8: Assessing risk of bias in included studies. In *Cochrane Handbook for systematic Reviews of Interventions (5.1.0)*.

Hingson, R. W., Zha, W., & Weitzman, E. R. (2009). 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. Supplement*, (16), 12–20. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2701090&tool=pmcentrez&rendertype=abstract

Holder, H. D., Longabaugh, R., Miller, W. R., & Rubonis, A. V. (1991). The cost effectiveness of treatment for alcohol problems: a first approximation. *Journal of Studies on Alcohol*, *52*, 517–540.

Hollis, R., & Campbell, F. (1999). What is meant by intention to treat analysis? *British Medical Journal*, *319*(7211), 670–674.

Ibrahim, J., Chen, M., Lipsitz, S., & Herring, A. (2005). Missing-data methods for generalized linear models. *Journal of American Statistical Association*, *100*(469), 332–346.

Jüni, P., Altman, D., & Egger, M. (2001). Systematic Reviews In Health Care : Assessing The Quality Of Controlled Clinical Trials. *BMJ: British Medical Journal*, *323*(7303), 42–46. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1120670/

Kaner, E., Beyer, F., Dickinson, H., Burnand, B., Pienaar, E., Saunders, J., … Campbell, F. (2007). Effectiveness of brief alcohol interventions in primary care populations. *Cochrane Database of Systematic Reviews*.

Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A Comparison of Web and Mail Survey Response Rates. *Public Opinion Quarterly*, *68*(1), 94–101. doi:10.1093/poq/nfh006

Kjaergard, L. (2001). Reported Methodologic Quality and Discrepancies between Large and and small randomized trials in meta-analyses. *Annals of Internal Medicine*. Retrieved from http://annals.org/article.aspx?articleid=714939

Kypri, K. (2003). Acceptability of Various Brief Intervention Approaches for Hazardous Drinking Among University Students. *Alcohol and Alcoholism*, *38*(6), 626–628. doi:10.1093/alcalc/agg121

Larimer, M. E., & Cronce, J. M. (2007). Identification, prevention, and treatment revisited: individual-focused college drinking prevention strategies 1999-2006. *Addictive Behaviors*, *32*(11), 2439–68. doi:10.1016/j.addbeh.2007.05.006

Lewis, M. A., & Neighbors, C. (2006). Social Norms Approaches Using Descriptive Drinking Norms Education: A Reviw of the Research on Personalized Normative Feedback. *Journal of American College Health*, *54*(1), 213–218. Retrieved from http://www.tandfonline.com/doi/abs/10.3200/JACH.54.4.213-218

Luckie, L. F., White, R. E., Miller, W. R., Icenogle, M. V., & Lasoski, M. C. (1992). Prevalence of alcohol problems in a V.A. outpatient population.

Moore, M. H., & Gerstein, D. R. (1981). *Alcohol and Public Policy: Beyond the Shadow of Prohibition*. (National Academy Press, Ed.). Washington, D.C.

Moreira, M., Smith, L., & Foxcroft, D. (2010). Social norms interventions to reduce alcohol misuse in University or College students. *The Cochrane Library*, (1), 1–76. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD006748.pub2/pdf/standard

Perkins, H. W. (2007). Misperceptions of peer drinking norms in Canada: another look at the "reign of error" and its consequences among college students. *Addictive Behaviors*, *32*(11), 2645–56. doi:10.1016/j.addbeh.2007.07.007

Perkins, H., & Wechsler, H. (1996). Variation in perceived college drinking norms and its impact on alcohol abuse: A nationwide study. *Journal of Drug Issues*, *26*(4), 961–974. Retrieved from http://psycnet.apa.org/psycinfo/1997-07672-013

Riper, H., van Straten, A., Keuken, M., Smit, F., Schippers, G., & Cuijpers, P. (2009). Curbing Problem Drinking with Personalized-Feedback Interventions. *American Journal of Preventive Medicine*, *36*(3), 247–255. doi:10.1016/j.amepre.2008.10.016

Schulenberg, J. E., & Maggs, J. L. (2002). A Developmental Perspective on Alcohol Use and Heavy Drinking during Adolescence and the Transition to Young Adulthood. *Journal of Studies on Alcohol. Supplement*, (14), 54–70. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12022730

Schulz, K., Chalmers, I., Hayes, R., & Altman, D. (1995). Empirical Evidence of Bias. *JAMA: The Journal of the …*, (273), 408–412. Retrieved from http://jama.ama-assn.org/content/273/5/408.short

Schulz, K. F. (1995). Unbiased research and the human spirit: the challenges of randomized controlled trials. *CMAJ : Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne*, *153*(6), 783–6. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1487264&tool=pmcentrez&rendertype=abstract

Smit, F., Cuijpers, P., Oostenbrink, J., Batelann, N., de Graaf, R., & Beekman, A. (2006). Costs of nine common mental disorders: implications for curative and preventive psychiatry. *Lournal of Mental Health Policy and Economics*, *9*, 193–200.

Stukel, T., & Fisher, E. (2007). Analysis of Observational Studies in the presence of treament selection bias. *JAMA: The Journal of …*, *297*(3). Retrieved from http://jama.ama-assn.org/content/297/3/278.short

Substance Abuse and Mental Health Services Administration. (2006). *Results from the 2006 National Survey on Drug Use and Health: National Findings*. Rockville, MD.

Turner, R. M., Spiegelhalter, D. J., Smith, G. C. S., & Thompson, S. G. (2009). Bias modelling in evidence synthesis. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, *172*(1), 21–47. doi:10.1111/j.1467-985X.2008.00547.x

Wechsler, H., Lee, J. E., Kuo, M., Seibring, M., Nelson, T. F., & Lee, H. (2001). Trends in College Binge Drinking During a Period of Increased Prevention Efforts. *Journal of American College Health*, *50*(5), 203–217.

White, H., & Jackson, K. (2004). Social and psychological influences on emerging adult drinking behavior. *Alcohol Research and Health*, *28*(4), 182–190. Retrieved from http://psycnet.apa.org/psycinfo/2006-01535-002

Wooldridge, J. (2002). *Econometric Analysis of Cross Section Panel Data*. Cambridge, MA: MIT press.

Appendix

Section A: Characteristics of Studies

All study characteristics for the studies analyzed in the previous Cochrane review can be found in the appendix of Moreira et al. (2010).

Agostinelli, Brown, and Miller (1995)

Methods	Design: RCT Follow-up: 6 weeks Attrition: 12%
Participants	Age: Unreported % Female: 48% Size: N =26 Setting: University Country: USA
Interventions	Program type: BMI Type: Mailed feedback Theoretical Base: Key components: personalized individual normative feedback Duration: Primary Staff: blind research assistants Control Group: Nothing (given same materials at end of study) Normative feedback: Received score based on past 60 days consumption; norms used were gender based and converted into a percentile score; feedback on risk of alcohol problems: calculated from tolerance, and family history; estimated peak BAC provided
Outcomes	List: for past 6 weeks: total alcohol consumption, average BAC, peak BAC in period Found a significant decrease in alcohol consumption as a result of feedback and a significant decrease in average weekly BAC as a result of feedback

Bendtsen et al. (2012)

Methods	Design: RCT Follow-up: 2 months Attrition: 48% control; 59% treatment
Participants	Age: Unreported % Female: 47% Size: N = 5536

	Setting: University Country: Sweden
Interventions	Program type: BMI Type: web feedback Theoretical Base: Key components: online survey of consumption followed by personalized normative feedback Duration: Primary Staff: Control Group: only contacted for post-treatment for consumption, additional pretreatment control condition that received pretreatment survey as well as post treatment Normative feedback: based on e-SBI3 statements summarizing weekly consumption, frequency of heavy episodic drinking, and highest BAC in past 3 months. Compared these values against safe drinking limits established by Swedish National Institute. Then given comparison of Swedish University student consumption. Additionally they gave people, if applicable, personalized advice for how to decrease unhealthy consumption. Problem drinker were >= 8 for men and >= 6 for women summing items 7-10 and 4-6 on AUDIT scale.
Outcomes	List: AUDIT score: total, dependency, problem score, dependence score, weekly consumption (g). Additional measures just for problem drinkers (weekly consumption, absolute change in consumption, and relative change). No effects found

Bewick, Trusler, Mulhern, Barkham, Hill (2008)

Methods	Design: RCT Follow-up: Attrition: 37%
Participants	Age: 21.29 (SD = 3.68) % Female: 69 % Size: n = 506 (completed pre-study assessment) Setting: University Country: UK
Interventions	Program type: BMI Type: website, contacted by email Theoretical Base: Key components: personalized normative feedback, generic information, medically recommended consumption

	Duration: Pre-survey data collected at 1 week, additional invitation to website at 6 weeks for intervention, 12 weeks post survey data collected Primary Staff: Control Group: AO Normative feedback: presented with amount of alcohol that participant consumed in a week and associated health risks; statements were made whether consumption should be reduced or maintained; recommended the number of days to consume no alcohol; given statements about what percentage of students consume less alcohol than themselves; negative effects reported from students in the same risk category were presented to participants; finally participants were given information on how to calculate units of consumption and general health risks of consuming lots of alcohol; tips for sensible drinking were provided
Outcomes	List: alcohol consumption per occasion significantly decreased for intervention condition; no decrease on CAGE score or number of drinks per week They found a significant reduction in average consumption per occasion.

Bewick et al. (2013)

Methods	Design: RCT Follow-up: 34 weeks Attrition: 1 Week (35%), 16 Weeks (54%), and 34 Weeks (60%)
Participants	Age: 17-50 mean = 20.8 % Female: 69% Size: n = 1478 at T0 (decided to participate) Setting: University Country: England
Interventions	Program type: BMI Type: Website feedback Theoretical Base: Key components: personalized feedback and social norms information; intervention participants were allowed to visit the feedback site between T1 and T2 (15 weeks) Duration: initial measurement of consumption at T0, T1 (week 1), T2 (week 16), T3 (week 34) Primary Staff: Control Group: assessment only Normative feedback: presented with amount of alcohol that

	participant consumed in a week and associated health risks; statements were made whether consumption should be reduced or maintained; recommended the number of days to consume no alcohol; given statements about what percentage of students consume less alcohol than themselves; negative effects reported from students in the same risk category were presented to participants; finally participants were given information on how to calculate units of consumption and general health risks of consuming lots of alcohol; tips for sensible drinking were provided
Outcomes	Units consumed in the last week had a significant reduction. CAGE scores and units in an average drinking occasion did not decrease significantly.

Cucciare et al. (2013)

Methods	Design: RCT Follow-up: 3 and 6 month Attrition: 15%
Participants	Age: mean = 59, (SD =15) % Female: 12% Size: n = 167 Setting: Veterans (screened for alcohol misuse during primary care visit) Country: USA
Interventions	Program type: BMI Type: web Theoretical Base: Key components: treatment as usual, personalized normative feedback Duration: 10-15 minutes Primary Staff: Control Group: Treatment as usual Normative feedback: summary of weekly consumption (alcohol and others), gender and age matched normative feedback on alcohol use in the population, summary of financial/social/health consequences of misusing alcohol, education on tolerance and peak BAC, summary of risk factors for unsafe drinking, self-reported motivation to change habits
Outcomes	No differences on between group means. Differences are apparent when you examine paired t-tests. Three and six month scores are the same. The found a general decrease in

	consumption for both groups over time (all 4 outcome variables)
Notes	This is an edge case for included studies. Even though the control is active I include it because the difference between treatments is just PNF. Inclusion into the study was based in AUDIT-C scores (>= 4 for men and >=3 for women). Confirmed that all patients had treatment as usual (TAU) at least 2 weeks before. Treatment as usual contained information on: typical alcohol consumption, lifetime negative consequences of alcohol and other substance use, risk factors for unsafe drinking, lifetime use of illicit substances (other than alcohol), motivation and confidence to change substance use.

Hansen et al. (2012)

Methods	Design: RCT Follow-up: 6 and 12 months Attrition: 37% at 6 months and 23% at 12 months
Participants	Age: 58 % Female: 45% Size: n = 1380 Setting: University Country: Adults (heavy drinkers)
Interventions	Program type: BMI Type: web Theoretical Base: Key components: Control Condition, Web brief advice, web personalized normative feedback Duration: Primary Staff: Control Group: had AO control and Brief advice all the same info as PNF, except PNF Normative feedback: summary of weekly consumption, comparison of weekly to maximum limit, and comparison to average level in municipality (gender specific). Also had info on health risks of drinking, risks to social relationships, and links for additional self-help and treatment centers
Outcomes	PNF was not effective at reducing the amount of drinking. Overall alcohol consumption goes down in all groups.

Kypri et al. (2009)

Methods	Design: RCT Follow-up: 1 month and 6 months Attrition: 35% at 6 months, missing at 1 month = 21.8%
Participants	Age: control = 19.7, intervention = 19.7 % Female: control =45.5%, intervention = 45.1% Size: control = 1184, intervention = 1251 Setting: University Country: Australia
Interventions	Program type: BMI Type: web Theoretical Base: AUDIT score with explanation of health risks, estimated highest BAC, estimates of expenditures, comparison of episodic and weekly drinking to others students (matched on age and gender), web resources for helping reduce drinking Key components: AUDIT score with explanation of health risks, estimated highest BAC, estimates of expenditures, comparison of episodic and weekly drinking to others students (matched on age and gender), web resources for helping reduce drinking Duration: Primary Staff: Control Group: Assessment only Normative feedback: bar graph of episodic and weekly consumption matched on age and gender, in 1 month follow-up students were also compared to their answers at baseline
Outcomes	At 1 month intervention group lower frequency of drinking, fewer drinks per occasion, and lower total consumption. No differences on log(APS) scores or AREAS score. At 6 months only differences in frequency and total consumption remained significant.

Martens et al. (2013)

Methods	Design: RCT Follow-up: 1 month and 6 month Attrition: at 1 month = 3.9%, at 6 months 5.5%
Participants	Age: 20.10 (SD = 1.35)

	% Female: 65.2% Size: n = 121+133 Setting: University (at least one binge drinking episode) Country: USA
Interventions	Program type: MI-based framework Type: Individual interview Theoretical Base: Key components: PNF condition: PNF with 4 reference groups (detailed below); protective behavioral strategies Duration: 15-20 min Primary Staff: graduate students in counseling or clinical psychology Control Group: Alcohol education: educational information about harmful effects of alcohol Normative feedback: Average drinks per week, average drinks per day, for typical male college student nationwide, typical female college student nationwide, typical male student at the university individual is attending, typical female student at the university that the student is attending; given a handout of self-reported alcohol use, perceptions of alcohol use (male and female), actual use of male and females, given percentile rank based on drinks per week,
Outcomes	List: Alcohol consumption (average drinks per week, average number of drinking days per week, peak BAC) used modified DDQ. Asked about consumption in past 30 days. Alcohol related problems via RAPI; descriptive norms via Drinking Norms Rating Form (just like DDQ but asking about a particular reference group's consumption; PNF decreases drinks per week, drinking days per week, and peak BAC at both time periods. Alcohol related problems did not decrease for PNF. It turns out that they did not consider alcohol education to be a reasonable control (good assumption), which means they never compared the groups so the statistics they have are subject to historical biases.

Moreira, Oskrochi, Foxcroft (2012)

Methods	Design: RCT (3 solomon group design) Follow-up: 6 and 12 months Attrition: at 6 months (PNF = 49%, Control = 50.5%), at 12 months (PNF = 40%, Control = 58%)

Participants	Age: Unreported % Female: 61% Size: 2611 Setting: University Country: UK
Interventions	Program type: BMI Type: Web Theoretical Base: Key components: Intervention had PNF, how much money is spent on alcohol, what is considered risky drinking, AUDIT score categories, how quickly alcohol is metabolized, sensible drinking Duration: Primary Staff: Control Group: main control (survey with alcohol questions at baseline), secondary control (only demographic questions at baseline) Normative feedback: drinking behavior assessments compared with average levels of drinking in student peer group. Included average number of drinks per week, days per week drinking,
Outcomes	List: AUDIT, frequency of alcohol consumption, quantity of alcohol consumption, alcohol related problems, adapted Drinking Norms Rating Form (DNFR), Alcohol Expectancies Questionnaire (AEQ-A) No effects found of the intervention. Similar results were found on just the high-risk drinkers as well.

Neighbors et al. (2004)

Methods	Design: RCT Follow-up: 3, 6 months Attrition: 21.4% at 3 month, 17.9% at 6 month
Participants	Age: 18.5 % Female: 59% Size: 252 Setting: University Country: USA
Interventions	Program type: BMI Type: Computer Theoretical Base: Key components: PNF on computer

	Duration: Primary Staff: Control Group: Assessment only Normative feedback: modeled on BASICS; summary of participants' perceived norms compared to actual norms and individuals behavior; given percentile rank relative to population
Outcomes	List: Drinking Norms Rating Form, overall consumption, peak quantity, typical weekly drinking, alcohol-related problems, Daily Drinking Questionnaire, Social Reasons for drinking (from Social Rewards sub-scale of Drinking Motives Questionnaire) It corrects misperceptions reasonably well. As for drinking behavior it is only moderately effective among heavy drinkers. The effects of pure PNF are somewhat less than more complex interventions that involve additional components.

Terlecki, Larimer, Copeland (2010)

Methods	Design: RCT Follow-up: 4 weeks Attrition: 8.7%
Participants	Age: 18-24 % Female: 38% Size: n = 92 Setting: University (Mandated Students) Country: USA
Interventions	Program type: BMI Type: Individual interview Theoretical Base: Key components: BASICS Duration: 50 min (baseline) Primary Staff: Graduate students in Clinical Psychology Control Group: Normative feedback: graphic of feedback of typical drinking patterns and perceived norms, percentile rank relative to campus norms
Outcomes	List: AUDIT, RAPI (Rutgers Alcohol Problem Index), DDQ (Daily Drinking Questionnaire)

	Total drinking quantity was significantly lowered.

Voogt et al. (2012)

Methods	Design: RCT Follow-up: 1 and 6 months Attrition: 6.3% (1 month), 8.7% (6 months)
Participants	Age: 18-24 (Mean = 20.08, SD = 1.7) % Female: 39.8% Size: 913 Setting: University Country: Netherlands
Interventions	Program type: BMI Type: Web Theoretical Base: Key components: What do you drink: personal drinking profile, risk factor, and normative comparisons; teaches individuals about action plans to consume less; giving tips on how to refuse consumption in social situations Duration: 20 min Primary Staff: Control Group: Assessment only Normative feedback: weekly consumption relative to same gender peer group; it was also tailored by alcohol intake and perceived social norms
Outcomes	List: Heavy Drinking, frequency of binge drinking, weekly consumption, No differences were found in heavy drinking or frequency of drinking by condition. Weekly alcohol consumption was the same for each treatment group.
Notes	In this case it appears that heavy drinking and frequency of binge drinking are measuring the same construct. Since heavy drinking has better face validity it was used in the meta-analysis

White et al. (2008)

Methods	Design: RCT Follow-up: 2 months Attrition: 13.5%

Participants	Age: Primarily 1st and 2nd year university students % Female: 28.7% Size: n = 230 Setting: University Country: USA
Interventions	Program type: BMI Type: Paper printout Theoretical Base: Key components: Personalized feedback profile, typical and heaviest peak BAC, alcohol problems, alcohol expectancies, risky behavior, and personal risk factors; educational information about effects of alcohol Duration: Primary Staff: Control Group: Delayed treatment Normative feedback: Comparison of drinking with college students of the same gender;
Outcomes	List: frequency of binge drinking, monthly frequency, peak BAC per day, RAPI, Social Desirability scale (shortened), Found no effects were found of condition on consumption levels

Section B: Adjustment Factors

Study	Adjustment Level Full	Adjustment Level Moderate
Agostinelli (1995)	1.90	1.62
Bendtsen (2012)	0.93	0.92
Bewick (2008)	1.08	1.05
Bewick (2013)	1.33	1.25
Borsari (2000)	1.54	1.36
Borsari (2005)	1.54	1.36
Carey (2006)	1.90	1.62
Collins (2002)	1.90	1.62
Cucciare (2013)	1.08	1.05
Hansen (2012)	1.44	1.36
Juarez (2006)	1.90	1.62
Kypri (2004)	1.08	1.05
Kypri (2005)	0.93	0.92
Kypri (2008)	0.93	0.92
Kypri (2009)	1.00	1.00
Lewis (2007a)	1.90	1.62
Lewis (2007b)	1.90	1.62
Marlatt (1998)	1.54	1.36
Martens (2013)	1.54	1.36
McNally (2003)	1.54	1.36
Michael (2006)	2.05	1.76
Moreira (2012)	1.16	1.14
Murphy (2001)	2.05	1.76
Neal (2004)	2.05	1.76
Neighbors (2004)	2.05	1.76
Neighbors (2006)	1.90	1.62
Terlecki (2010)	1.90	1.62
Voogt (2013)	1.66	1.48
Walters (2000)	1.90	1.62
Walters (2007)	2.05	1.76
Werch (2000)	1.90	1.62
White (2008)	1.66	1.48

Table 8: Total adjustment factor applied to each study

Section C: Additional Forest Plots

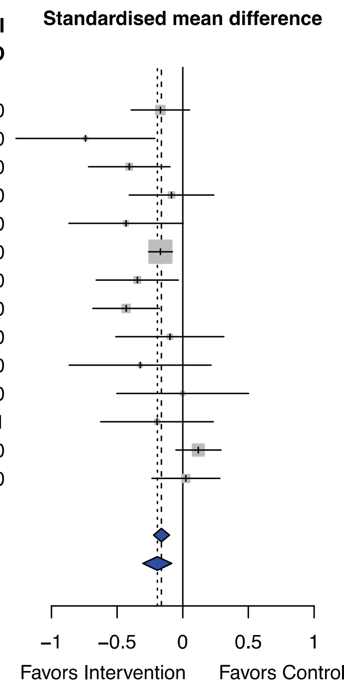
Short-Term Follow-up (0-3 months)

Frequency of Consumption

Forest Plot of Frequency of Consumption Measured at 0–3 Months,
No Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%–CI
Bewick (2008)	138	12.020	13.580	179	14.850	18.670	−0.17	[−0.39; 0.05]
Borsari (2000)	29	3.830	0.890	30	4.570	1.070	−0.74	[−1.27; −0.21]
Carey (2006)	84	4.400	2.100	79	5.300	2.300	−0.41	[−0.72; −0.10]
Cucciare (2013)	82	17.800	10.500	68	18.700	10.500	−0.09	[−0.41; 0.24]
Kypri (2004)	42	3.170	1.770	41	4.120	2.530	−0.43	[−0.87; 0.00]
Kypri (2009)	962	−0.117	0.690	942	0.000	0.690	−0.17	[−0.26; −0.08]
Lewis (2007b)	76	3.430	1.310	84	3.880	1.280	−0.35	[−0.66; −0.03]
Martens (2013)	116	2.310	1.350	128	2.880	1.290	−0.43	[−0.69; −0.18]
Michael (2006)	47	5.300	4.700	44	5.800	5.500	−0.10	[−0.51; 0.31]
Murphy (2001)	30	3.410	1.130	24	3.760	0.980	−0.32	[−0.86; 0.22]
Neal (2004)	31	2.100	1.400	30	2.100	1.500	0.00	[−0.50; 0.50]
Terlecki (2010)	41	3.169	1.563	43	3.496	1.741	−0.20	[−0.62; 0.23]
Werch (2000)	266	2.500	2.700	255	2.200	2.300	0.12	[−0.05; 0.29]
White (2008)	111	−0.300	1.290	119	−0.330	1.180	0.02	[−0.23; 0.28]
Fixed effect model	**2055**			**2066**			**−0.16**	**[−0.22; −0.10]**
Random effects model							**−0.19**	**[−0.30; −0.08]**

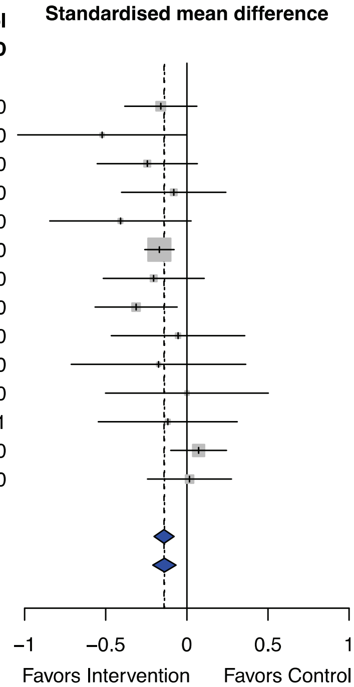
Heterogeneity: I−squared=52.7%, tau−squared=0.0189, p=0.0107



Forest Plot of Frequency of Consumption Measured at 0–3 Months, Moderate Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%–CI
Bewick (2008)	138	12.1516781	13.6823803	179	14.850	18.670	–0.16	[–0.38; 0.06]
Borsari (2000)	29	4.0272552	0.9775637	30	4.570	1.070	–0.52	[–1.04; 0.00]
Carey (2006)	84	4.7452983	2.2264232	79	5.300	2.300	–0.24	[–0.55; 0.06]
Cucciare (2013)	82	17.8418764	10.6701565	68	18.700	10.500	–0.08	[–0.40; 0.24]
Kypri (2004)	42	3.2142029	1.8016839	41	4.120	2.530	–0.41	[–0.84; 0.03]
Kypri (2009)	962	–0.1170000	0.6900000	942	0.000	0.690	–0.17	[–0.26; –0.08]
Lewis (2007b)	76	3.6026491	1.4193623	84	3.880	1.280	–0.20	[–0.52; 0.11]
Martens (2013)	116	2.4619398	1.3777837	128	2.880	1.290	–0.31	[–0.57; –0.06]
Michael (2006)	47	5.5164858	4.7822479	44	5.800	5.500	–0.05	[–0.47; 0.36]
Murphy (2001)	30	3.5615400	1.2182833	24	3.760	0.980	–0.17	[–0.71; 0.36]
Neal (2004)	31	2.1000000	1.4333835	30	2.100	1.500	0.00	[–0.50; 0.50]
Terlecki (2010)	41	3.2944584	1.6471329	43	3.496	1.741	–0.12	[–0.55; 0.31]
Werch (2000)	266	2.3849006	2.7483822	255	2.200	2.300	0.07	[–0.10; 0.24]
White (2008)	111	–0.3097571	1.2928672	119	–0.330	1.180	0.02	[–0.24; 0.27]
Fixed effect model	**2055**			**2066**			**–0.14**	**[–0.20; –0.08]**
Random effects model							**–0.14**	**[–0.21; –0.07]**

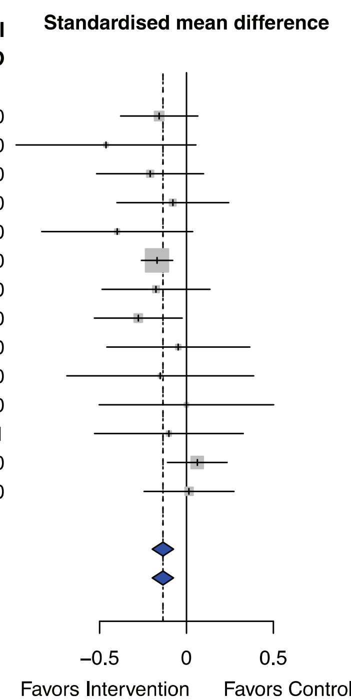
Heterogeneity: I–squared=9.2%, tau–squared=0.0017, p=0.3524



Forest Plot of Frequency of Consumption Measured at 0–3 Months, Full Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%–CI
Bewick (2008)	138	12.221496	13.6806320	179	14.850	18.670	–0.16	[–0.38; 0.07]
Borsari (2000)	29	4.088882	0.9764929	30	4.570	1.070	–0.46	[–0.98; 0.05]
Carey (2006)	84	4.826033	2.2227509	79	5.300	2.300	–0.21	[–0.52; 0.10]
Cucciare (2013)	82	17.864080	10.6667221	68	18.700	10.500	–0.08	[–0.40; 0.24]
Kypri (2004)	42	3.237640	1.8013331	41	4.120	2.530	–0.40	[–0.83; 0.04]
Kypri (2009)	962	–0.117000	0.6900000	942	0.000	0.690	–0.17	[–0.26; –0.08]
Lewis (2007b)	76	3.643017	1.4153437	84	3.880	1.280	–0.18	[–0.49; 0.14]
Martens (2013)	116	2.509409	1.3774457	128	2.880	1.290	–0.28	[–0.53; –0.02]
Michael (2006)	47	5.556190	4.7769417	44	5.800	5.500	–0.05	[–0.46; 0.36]
Murphy (2001)	30	3.589333	1.2132016	24	3.760	0.980	–0.15	[–0.69; 0.39]
Neal (2004)	31	2.100000	1.4309795	30	2.100	1.500	0.00	[–0.50; 0.50]
Terlecki (2010)	41	3.323792	1.6435902	43	3.496	1.741	–0.10	[–0.53; 0.33]
Werch (2000)	266	2.357989	2.7452052	255	2.200	2.300	0.06	[–0.11; 0.23]
White (2008)	111	–0.311940	1.2926491	119	–0.330	1.180	0.01	[–0.24; 0.27]
Fixed effect model	**2055**			**2066**			**–0.14**	**[–0.20; –0.07]**
Random effects model							**–0.14**	**[–0.20; –0.07]**

Heterogeneity: I–squared=0%, tau–squared=0, p=0.5269

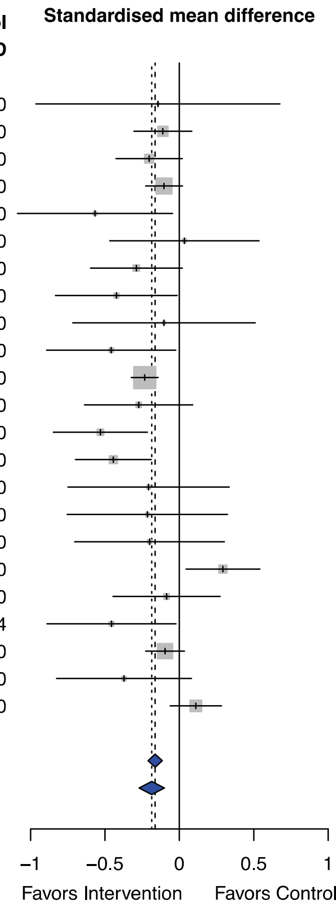


Average Consumption

Forest Plot of Average Consumption at 0–3 Months,
No Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%–CI
Agostinelli (1995)	12	8.500	7.200	11	10.100	13.600	–0.14	[–0.96; 0.68]
Bendtsen (2012)	201	131.400	107.300	207	143.300	107.300	–0.11	[–0.30; 0.08]
Bewick (2008)	138	8.460	5.680	179	9.800	7.220	–0.20	[–0.43; 0.02]
Bewick (2013)	457	16.200	16.200	544	18.000	18.500	–0.10	[–0.23; 0.02]
Borsari (2000)	29	11.400	7.030	30	15.780	8.170	–0.57	[–1.09; –0.05]
Borsari (2005)	31	18.100	11.960	30	17.710	10.490	0.03	[–0.47; 0.54]
Carey (2006)	84	13.700	9.500	79	16.400	9.100	–0.29	[–0.60; 0.02]
Collins (2002)	47	1.090	0.310	47	1.210	0.250	–0.42	[–0.83; –0.01]
Juarez (2006)	20	0.800	0.640	21	0.870	0.690	–0.10	[–0.72; 0.51]
Kypri (2004)	42	8.290	3.750	42	10.360	5.100	–0.46	[–0.89; –0.02]
Kypri (2009)	962	–0.186	0.800	942	0.000	0.800	–0.23	[–0.32; –0.14]
Lewis (2007a)	60	2.580	1.200	57	2.910	1.200	–0.27	[–0.64; 0.09]
Lewis (2007b)	76	14.780	6.710	84	18.350	6.690	–0.53	[–0.85; –0.21]
Martens (2013)	116	10.140	8.900	128	14.490	10.520	–0.44	[–0.70; –0.19]
McNally (2003)	24	6.760	7.540	29	8.150	5.790	–0.21	[–0.75; 0.34]
Murphy (2001)	30	17.580	7.810	24	19.490	9.840	–0.21	[–0.75; 0.32]
Neal (2004)	31	4.300	3.400	30	5.000	3.500	–0.20	[–0.70; 0.30]
Neighbors (2004)	126	12.140	9.205	126	9.450	9.090	0.29	[0.04; 0.54]
Neighbors (2006)	58	10.700	9.140	61	11.560	10.680	–0.09	[–0.45; 0.27]
Terlecki (2010)	41	14.078	8.791	43	18.863	11.714	–0.46	[–0.89; –0.02]
Voogt (2013)	456	28.600	22.600	451	31.000	26.900	–0.10	[–0.23; 0.03]
Walters (2007)	37	3.330	5.520	39	5.830	7.580	–0.37	[–0.83; 0.08]
Werch (2000)	266	2.900	2.900	255	2.600	2.500	0.11	[–0.06; 0.28]
Fixed effect model	**3344**			**3459**			**–0.16**	**[–0.21; –0.11]**
Random effects model							**–0.19**	**[–0.27; –0.10]**

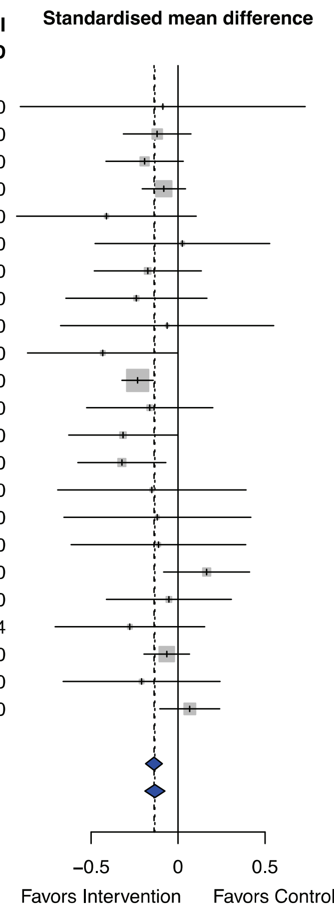
Heterogeneity: I–squared=53.4%, tau–squared=0.0173, p=0.0014



Forest Plot of Average Consumption at 0–3 Months, Moderate Adjustment of Effect Sizes

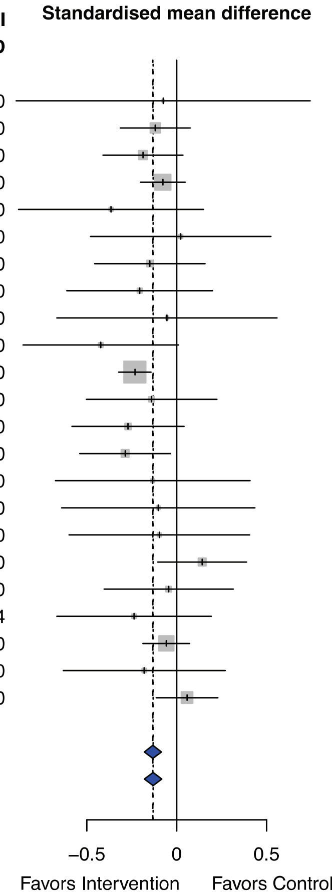
	Experimental			Control				
Study	**Total**	**Mean**	**SD**	**Total**	**Mean**	**SD**	**SMD**	**95%–CI**
Agostinelli (1995)	12	9.1138636	7.3878378	11	10.100	13.600	–0.09	[–0.91; 0.73]
Bendtsen (2012)	201	130.3652174	108.0127458	207	143.300	107.300	–0.12	[–0.31; 0.07]
Bewick (2008)	138	8.5223494	5.7570848	179	9.800	7.220	–0.19	[–0.41; 0.03]
Bewick (2013)	457	16.5577755	16.5131525	544	18.000	18.500	–0.08	[–0.21; 0.04]
Borsari (2000)	29	12.5675374	7.1703802	30	15.780	8.170	–0.41	[–0.93; 0.10]
Borsari (2005)	31	17.9960412	12.1461945	30	17.710	10.490	0.02	[–0.48; 0.53]
Carey (2006)	84	14.7358948	9.8231287	79	16.400	9.100	–0.17	[–0.48; 0.13]
Collins (2002)	47	1.1360398	0.3525145	47	1.210	0.250	–0.24	[–0.65; 0.17]
Juarez (2006)	20	0.8268565	0.6571202	21	0.870	0.690	–0.06	[–0.68; 0.55]
Kypri (2004)	42	8.3863158	3.8434128	42	10.360	5.100	–0.43	[–0.87; 0.00]
Kypri (2009)	962	–0.1860000	0.8000000	942	0.000	0.800	–0.23	[–0.32; –0.14]
Lewis (2007a)	60	2.7066094	1.2720260	57	2.910	1.200	–0.16	[–0.53; 0.20]
Lewis (2007b)	76	16.1496832	7.1579513	84	18.350	6.690	–0.32	[–0.63; 0.00]
Martens (2013)	116	11.2995406	9.0110428	128	14.490	10.520	–0.32	[–0.58; –0.07]
McNally (2003)	24	7.1305199	7.6083450	29	8.150	5.790	–0.15	[–0.69; 0.39]
Murphy (2001)	30	18.4069757	8.1812550	24	19.490	9.840	–0.12	[–0.66; 0.42]
Neal (2004)	31	4.6030801	3.4707753	30	5.000	3.500	–0.11	[–0.61; 0.39]
Neighbors (2004)	126	10.9753065	9.3690388	126	9.450	9.090	0.16	[–0.08; 0.41]
Neighbors (2006)	58	11.0299517	9.3652348	61	11.560	10.680	–0.05	[–0.41; 0.31]
Terlecki (2010)	41	15.9138358	9.1649039	43	18.863	11.714	–0.28	[–0.71; 0.15]
Voogt (2013)	456	29.3805668	22.7275404	451	31.000	26.900	–0.07	[–0.20; 0.07]
Walters (2007)	37	4.4124289	5.5867665	39	5.830	7.580	–0.21	[–0.66; 0.24]
Werch (2000)	266	2.7849006	2.9560714	255	2.600	2.500	0.07	[–0.10; 0.24]
Fixed effect model	**3344**			**3459**			**–0.14**	**[–0.19; –0.09]**
Random effects model							**–0.13**	**[–0.19; –0.08]**

Heterogeneity: I–squared=13.5%, tau–squared=0.0023, p=0.2772



Forest Plot of Average Consumption at 0–3 Months, Full Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%–CI
Agostinelli (1995)	12	9.2573926	7.3806943	11	10.100	13.600	–0.08	[–0.89; 0.74]
Bendtsen (2012)	201	130.4480000	108.0042127	207	143.300	107.300	–0.12	[–0.31; 0.08]
Bewick (2008)	138	8.5554080	5.7557915	179	9.800	7.220	–0.19	[–0.41; 0.04]
Bewick (2013)	457	16.6458096	16.4993888	544	18.000	18.500	–0.08	[–0.20; 0.05]
Borsari (2000)	29	12.9322992	7.1706616	30	15.780	8.170	–0.37	[–0.88; 0.15]
Borsari (2005)	31	17.9635624	12.1381746	30	17.710	10.490	0.02	[–0.48; 0.52]
Carey (2006)	84	14.9781001	9.8120450	79	16.400	9.100	–0.15	[–0.46; 0.16]
Collins (2002)	47	1.1468044	0.3509039	47	1.210	0.250	–0.21	[–0.61; 0.20]
Juarez (2006)	20	0.8331359	0.6563096	21	0.870	0.690	–0.05	[–0.67; 0.56]
Kypri (2004)	42	8.4373840	3.8422872	42	10.360	5.100	–0.42	[–0.85; 0.01]
Kypri (2009)	962	–0.1860000	0.8000000	942	0.000	0.800	–0.23	[–0.32; –0.14]
Lewis (2007a)	60	2.7362122	1.2692515	57	2.910	1.200	–0.14	[–0.50; 0.22]
Lewis (2007b)	76	16.4699323	7.1472153	84	18.350	6.690	–0.27	[–0.58; 0.04]
Martens (2013)	116	11.6618040	9.0107741	128	14.490	10.520	–0.29	[–0.54; –0.03]
McNally (2003)	24	7.2462776	7.6065640	29	8.150	5.790	–0.13	[–0.67; 0.41]
Murphy (2001)	30	18.5586458	8.1595296	24	19.490	9.840	–0.10	[–0.64; 0.43]
Neal (2004)	31	4.6586660	3.4666941	30	5.000	3.500	–0.10	[–0.60; 0.41]
Neighbors (2004)	126	10.7616978	9.3537773	126	9.450	9.090	0.14	[–0.11; 0.39]
Neighbors (2006)	58	11.1070985	9.3543056	61	11.560	10.680	–0.04	[–0.40; 0.31]
Terlecki (2010)	41	16.3430774	9.1585177	43	18.863	11.714	–0.24	[–0.67; 0.19]
Voogt (2013)	456	29.5552000	22.7186433	451	31.000	26.900	–0.06	[–0.19; 0.07]
Walters (2007)	37	4.6109500	5.5841416	39	5.830	7.580	–0.18	[–0.63; 0.27]
Werch (2000)	266	2.7579889	2.9524237	255	2.600	2.500	0.06	[–0.11; 0.23]
Fixed effect model	**3344**			**3459**			**–0.13**	**[–0.18; –0.08]**
Random effects model							**–0.13**	**[–0.18; –0.08]**



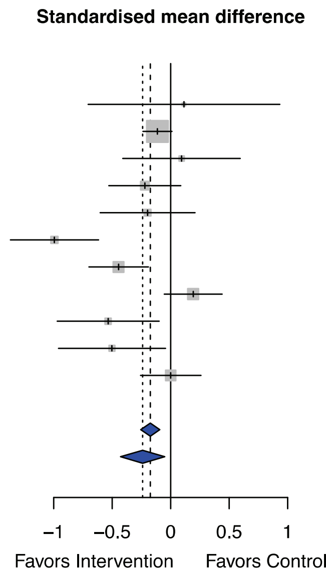
Heterogeneity: I–squared=1.1%, tau–squared=0.0002, p=0.4459

Peak BAC

Forest Plot of Peak BAC at 0–3 Months, No Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%–CI
Agostinelli (1995)	12	186.900	150.400	11	172.500	82.100	0.11	[–0.71; 0.93]
Bewick (2013)	457	9.820	7.130	544	10.640	7.260	–0.11	[–0.24; 0.01]
Borsari (2005)	31	0.170	0.090	30	0.160	0.120	0.09	[–0.41; 0.60]
Carey (2006)	84	0.160	0.090	79	0.180	0.090	–0.22	[–0.53; 0.09]
Collins (2002)	47	0.180	0.110	47	0.200	0.090	–0.20	[–0.60; 0.21]
Kypri (2005)	61	0.110	0.020	61	0.130	0.020	–0.99	[–1.37; –0.62]
Martens (2013)	116	0.107	0.097	128	0.152	0.104	–0.45	[–0.70; –0.19]
Neighbors (2004)	126	9.110	4.490	126	8.200	4.940	0.19	[–0.06; 0.44]
Terlecki (2010)	41	4.976	2.014	43	6.092	2.118	–0.53	[–0.97; –0.10]
Walters (2007)	37	0.050	0.090	39	0.110	0.140	–0.50	[–0.96; –0.04]
White (2008)	111	–0.010	0.040	119	–0.010	0.360	0.00	[–0.26; 0.26]
Fixed effect model	**1123**			**1227**			**–0.17**	**[–0.26; –0.09]**
Random effects model							**–0.24**	**[–0.43; –0.05]**

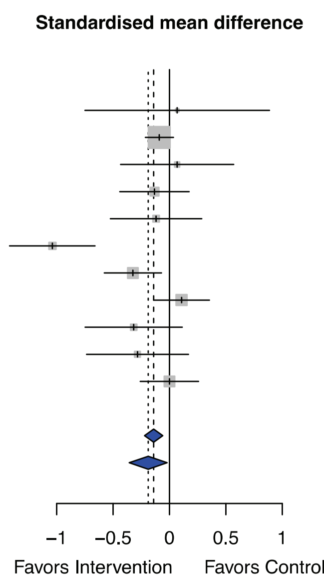
Heterogeneity: I–squared=74.9%, tau–squared=0.0669, p<0.0001



Forest Plot of Peak BAC at 0–3 Months, Moderate Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%–CI
Agostinelli (1995)	12	181.37522761	154.10330062	11	172.500	82.100	0.07	[–0.75; 0.89]
Bewick (2013)	457	9.98298662	7.32871274	544	10.640	7.260	–0.09	[–0.21; 0.03]
Borsari (2005)	31	0.16733439	0.09190782	30	0.160	0.120	0.07	[–0.43; 0.57]
Carey (2006)	84	0.16767329	0.09399781	79	0.180	0.090	–0.13	[–0.44; 0.17]
Collins (2002)	47	0.18767329	0.11428557	47	0.200	0.090	–0.12	[–0.52; 0.29]
Kypri (2005)	61	0.10826087	0.02156955	61	0.130	0.020	–1.04	[–1.42; –0.66]
Martens (2013)	116	0.11899525	0.09816096	128	0.152	0.104	–0.32	[–0.58; –0.07]
Neighbors (2004)	126	8.71599587	4.64623696	126	8.200	4.940	0.11	[–0.14; 0.35]
Terlecki (2010)	41	5.40416986	2.17527019	43	6.092	2.118	–0.32	[–0.75; 0.11]
Walters (2007)	37	0.07597829	0.09113724	39	0.110	0.140	–0.28	[–0.74; 0.17]
White (2008)	111	–0.01000000	0.04008931	119	–0.010	0.360	0.00	[–0.26; 0.26]
Fixed effect model	**1123**			**1227**			**–0.14**	**[–0.22; –0.06]**
Random effects model							**–0.19**	**[–0.36; –0.02]**

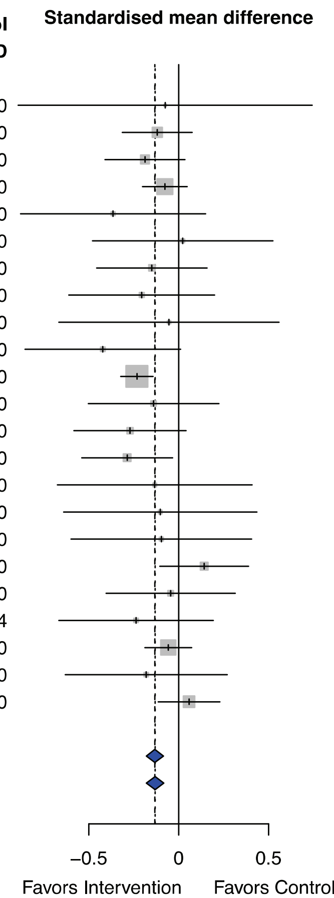
Heterogeneity: I–squared=68%, tau–squared=0.0474, p=0.0005



Forest Plot of Average Consumption at 0–3 Months,
Full Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%–CI
Agostinelli (1995)	12	9.2573926	7.3806943	11	10.100	13.600	−0.08	[−0.89; 0.74]
Bendtsen (2012)	201	130.4480000	108.0042127	207	143.300	107.300	−0.12	[−0.31; 0.08]
Bewick (2008)	138	8.5554080	5.7557915	179	9.800	7.220	−0.19	[−0.41; 0.04]
Bewick (2013)	457	16.6458096	16.4993888	544	18.000	18.500	−0.08	[−0.20; 0.05]
Borsari (2000)	29	12.9322992	7.1706616	30	15.780	8.170	−0.37	[−0.88; 0.15]
Borsari (2005)	31	17.9635624	12.1381746	30	17.710	10.490	0.02	[−0.48; 0.52]
Carey (2006)	84	14.9781001	9.8120450	79	16.400	9.100	−0.15	[−0.46; 0.16]
Collins (2002)	47	1.1468044	0.3509039	47	1.210	0.250	−0.21	[−0.61; 0.20]
Juarez (2006)	20	0.8331359	0.6563096	21	0.870	0.690	−0.05	[−0.67; 0.56]
Kypri (2004)	42	8.4373840	3.8422872	42	10.360	5.100	−0.42	[−0.85; 0.01]
Kypri (2009)	962	−0.1860000	0.8000000	942	0.000	0.800	−0.23	[−0.32; −0.14]
Lewis (2007a)	60	2.7362122	1.2692515	57	2.910	1.200	−0.14	[−0.50; 0.22]
Lewis (2007b)	76	16.4699323	7.1472153	84	18.350	6.690	−0.27	[−0.58; 0.04]
Martens (2013)	116	11.6618040	9.0107741	128	14.490	10.520	−0.29	[−0.54; −0.03]
McNally (2003)	24	7.2462776	7.6065640	29	8.150	5.790	−0.13	[−0.67; 0.41]
Murphy (2001)	30	18.5586458	8.1595296	24	19.490	9.840	−0.10	[−0.64; 0.43]
Neal (2004)	31	4.6586660	3.4666941	30	5.000	3.500	−0.10	[−0.60; 0.41]
Neighbors (2004)	126	10.7616978	9.3537773	126	9.450	9.090	0.14	[−0.11; 0.39]
Neighbors (2006)	58	11.1070985	9.3543056	61	11.560	10.680	−0.04	[−0.40; 0.31]
Terlecki (2010)	41	16.3430774	9.1585177	43	18.863	11.714	−0.24	[−0.67; 0.19]
Voogt (2013)	456	29.5552000	22.7186433	451	31.000	26.900	−0.06	[−0.19; 0.07]
Walters (2007)	37	4.6109500	5.5841416	39	5.830	7.580	−0.18	[−0.63; 0.27]
Werch (2000)	266	2.7579889	2.9524237	255	2.600	2.500	0.06	[−0.11; 0.23]
Fixed effect model	**3344**			**3459**			**−0.13**	**[−0.18; −0.08]**
Random effects model							**−0.13**	**[−0.18; −0.08]**

Heterogeneity: I−squared=1.1%, tau−squared=0.0002, p=0.4459

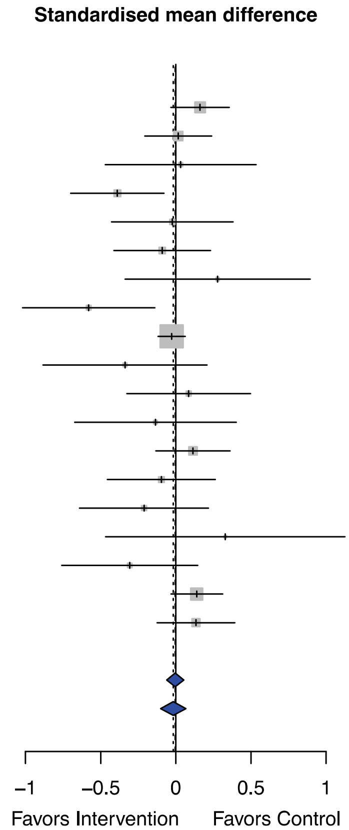


Alcohol Related Problems

Forest Plot of Alcohol Related Problems at 0–3 Months, No Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%–CI
Bendtsen (2012)	201	3.300	3.300	207	2.800	2.90	0.16	[–0.03; 0.36]
Bewick (2008)	138	1.570	1.110	179	1.550	1.36	0.02	[–0.21; 0.24]
Borsari (2005)	31	5.900	5.560	30	5.730	4.84	0.03	[–0.47; 0.53]
Carey (2006)	84	5.900	6.600	79	8.500	6.70	–0.39	[–0.70; –0.08]
Collins (2002)	47	7.830	6.670	47	7.980	5.69	–0.02	[–0.43; 0.38]
Cucciare (2013)	82	6.300	10.100	68	7.200	9.60	–0.09	[–0.41; 0.23]
Juarez (2006)	20	5.600	5.080	21	4.280	4.21	0.28	[–0.34; 0.89]
Kypri (2004)	42	2.360	1.820	41	3.540	2.20	–0.58	[–1.02; –0.14]
Kypri (2009)	962	–0.020	0.720	942	0.000	0.72	–0.03	[–0.12; 0.06]
McNally (2003)	24	4.250	4.270	29	5.890	5.16	–0.34	[–0.88; 0.21]
Michael (2006)	47	5.100	5.700	44	4.600	5.90	0.09	[–0.33; 0.50]
Murphy (2001)	30	7.230	3.810	24	7.780	4.19	–0.14	[–0.67; 0.40]
Neighbors (2004)	126	7.220	6.286	126	6.500	6.29	0.11	[–0.13; 0.36]
Neighbors (2006)	58	5.690	6.430	61	6.400	8.05	–0.10	[–0.46; 0.26]
Terlecki (2010)	41	8.997	9.165	43	10.902	8.68	–0.21	[–0.64; 0.22]
Walters (2000)	11	6.000	3.190	14	4.860	3.48	0.33	[–0.47; 1.12]
Walters (2007)	37	1.730	2.700	39	2.750	3.77	–0.31	[–0.76; 0.15]
Werch (2000)	266	2.700	4.000	255	2.200	3.10	0.14	[–0.03; 0.31]
White (2008)	111	–0.050	0.630	119	–0.130	0.56	0.13	[–0.12; 0.39]
Fixed effect model	**2358**			**2368**			**0.00**	**[–0.06; 0.05]**
Random effects model							**–0.02**	**[–0.10; 0.07]**

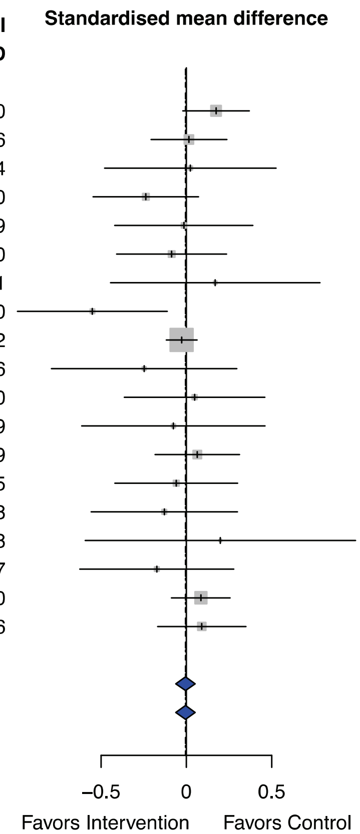
Heterogeneity: I–squared=32.6%, tau–squared=0.0092, p=0.0844



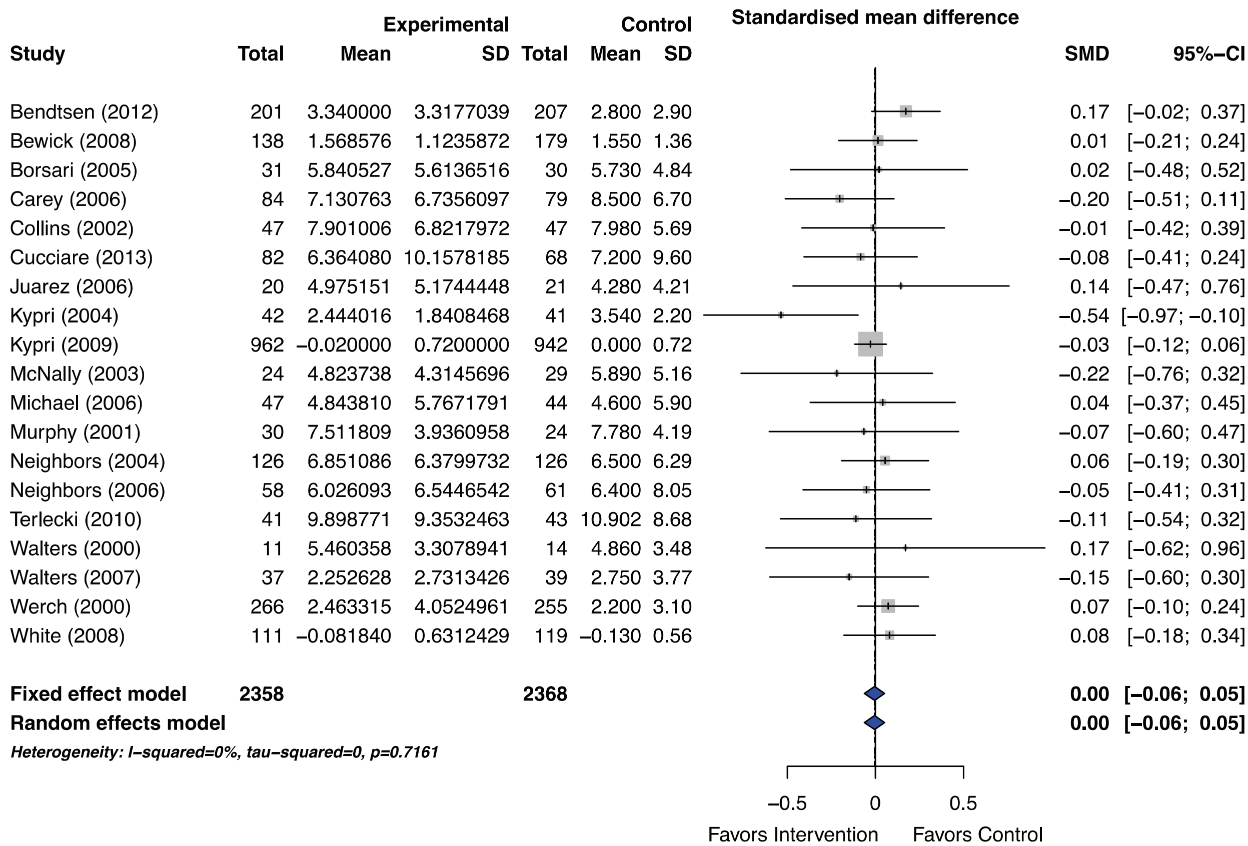
Forest Plot of Alcohol Related Problems at 0–3 Months,
Moderate Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%–CI
Bendtsen (2012)	201	3.34347826	3.3179510	207	2.800	2.90	0.17	[−0.02; 0.37]
Bewick (2008)	138	1.56906941	1.1238989	179	1.550	1.36	0.02	[−0.21; 0.24]
Borsari (2005)	31	5.85468462	5.6160752	30	5.730	4.84	0.02	[−0.48; 0.53]
Carey (2006)	84	6.89752835	6.7387801	79	8.500	6.70	−0.24	[−0.55; 0.07]
Collins (2002)	47	7.88754971	6.8305376	47	7.980	5.69	−0.01	[−0.42; 0.39]
Cucciare (2013)	82	6.34187643	10.1590078	68	7.200	9.60	−0.09	[−0.41; 0.24]
Juarez (2006)	20	5.09356253	5.1824285	21	4.280	4.21	0.17	[−0.44; 0.78]
Kypri (2004)	42	2.41490465	1.8409934	41	3.540	2.20	−0.55	[−0.99; −0.11]
Kypri (2009)	962	−0.02000000	0.7200000	942	0.000	0.72	−0.03	[−0.12; 0.06]
McNally (2003)	24	4.68716012	4.3150177	29	5.890	5.16	−0.25	[−0.79; 0.30]
Michael (2006)	47	4.88351422	5.7729174	44	4.600	5.90	0.05	[−0.36; 0.46]
Murphy (2001)	30	7.46813436	3.9445945	24	7.780	4.19	−0.08	[−0.61; 0.46]
Neighbors (2004)	126	6.90826047	6.3882074	126	6.500	6.29	0.06	[−0.18; 0.31]
Neighbors (2006)	58	5.96240197	6.5503281	61	6.400	8.05	−0.06	[−0.42; 0.30]
Terlecki (2010)	41	9.72788135	9.3608518	43	10.902	8.68	−0.13	[−0.56; 0.30]
Walters (2000)	11	5.56262219	3.3183292	14	4.860	3.48	0.20	[−0.59; 0.99]
Walters (2007)	37	2.17163100	2.7328149	39	2.750	3.77	−0.17	[−0.62; 0.28]
Werch (2000)	266	2.50816763	4.0562019	255	2.200	3.10	0.09	[−0.09; 0.26]
White (2008)	111	−0.07601889	0.6313432	119	−0.130	0.56	0.09	[−0.17; 0.35]
Fixed effect model	**2358**			**2368**			**0.00**	**[−0.06; 0.05]**
Random effects model							**0.00**	**[−0.06; 0.05]**

Heterogeneity: I–squared=0%, tau–squared=0, p=0.579

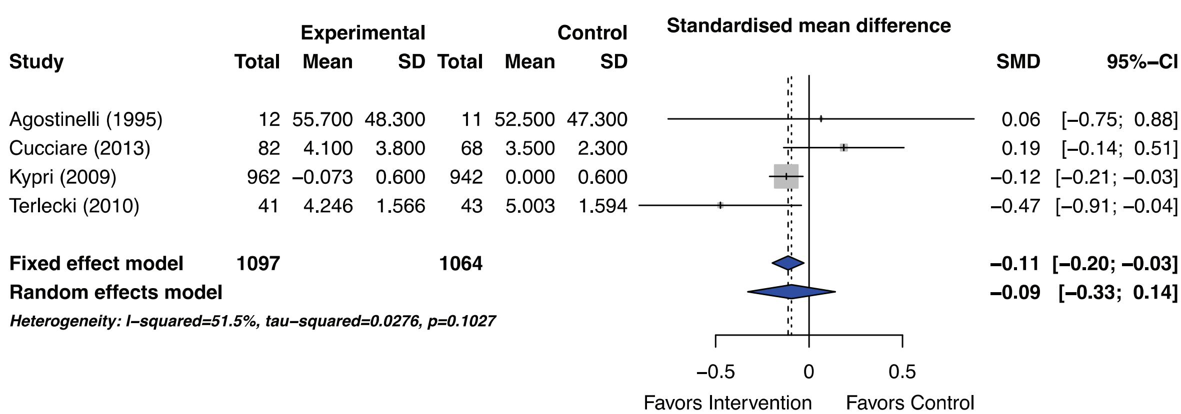


Forest Plot of Alcohol Related Problems at 0–3 Months,
Full Adjustment of Effect Sizes

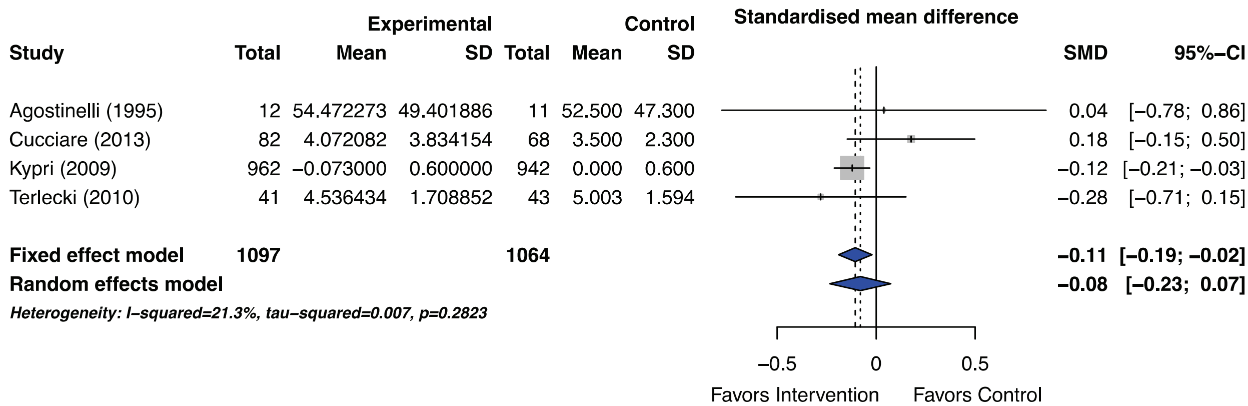


Average Consumption per Occasion

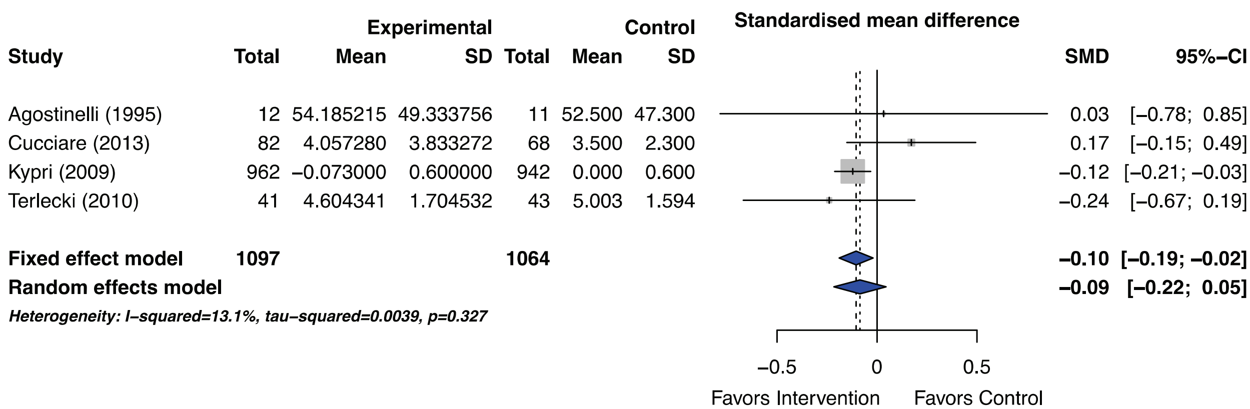
Forest Plot of Average Consumption Per Occasion at 0–3 Months,
No Adjustment of Effect Sizes



Forest Plot of Average Consumption Per Occasion at 0–3 Months,
Moderate Adjustment of Effect Sizes



Forest Plot of Average Consumption Per Occasion at 0–3 Months,
Full Adjustment of Effect Sizes

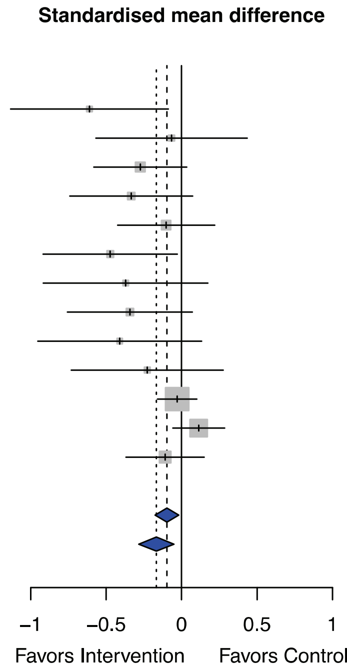


<u>Frequency of Heavy Drinking</u>

Forest Plot of Frequnecy of Heavy Drinking at 0–3 Months, No Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%–CI
Borsari (2000)	29	2.550	1.40	30	3.37	1.25	–0.61	[–1.13; –0.09]
Borsari (2005)	31	6.830	4.11	30	7.13	4.81	–0.07	[–0.57; 0.44]
Carey (2006)	84	5.100	4.00	79	6.20	4.00	–0.27	[–0.58; 0.03]
Collins (2002)	47	5.490	3.81	47	6.94	4.75	–0.33	[–0.74; 0.07]
Cucciare (2013)	82	10.000	19.00	68	12.00	20.00	–0.10	[–0.42; 0.22]
Kypri (2004)	40	1.230	1.46	40	2.08	2.05	–0.47	[–0.92; –0.03]
McNally (2003)	24	3.000	3.05	29	4.17	3.15	–0.37	[–0.92; 0.17]
Michael (2006)	47	2.700	3.20	44	4.20	5.30	–0.34	[–0.76; 0.07]
Murphy (2001)	30	1.970	1.07	24	2.45	1.25	–0.41	[–0.95; 0.13]
Neal (2004)	31	1.200	1.20	30	1.50	1.40	–0.23	[–0.73; 0.28]
Voogt (2013)	456	–0.083	2.79	451	0.00	2.79	–0.03	[–0.16; 0.10]
Werch (2000)	266	1.500	1.90	255	1.30	1.60	0.11	[–0.06; 0.29]
White (2008)	111	–0.170	0.67	119	–0.10	0.61	–0.11	[–0.37; 0.15]
Fixed effect model	**1278**			**1246**			**–0.10**	**[–0.18; –0.02]**
Random effects model							**–0.17**	**[–0.28; –0.05]**

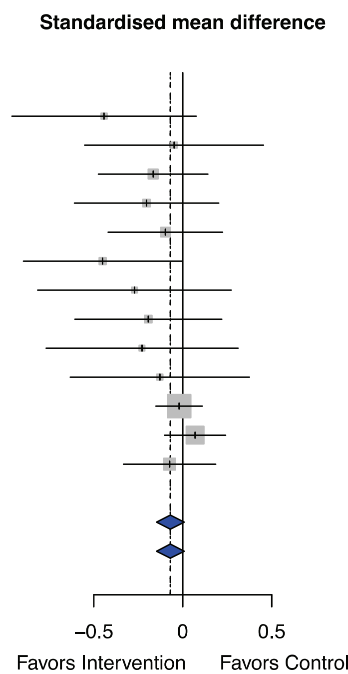
Heterogeneity: I–squared=39%, tau–squared=0.0152, p=0.0735



Forest Plot of Frequnecy of Heavy Drinking at 0–3 Months, Moderate Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%–CI
Borsari (2000)	29	2.7685801	1.432665	30	3.37	1.25	–0.44	[–0.96; 0.08]
Borsari (2005)	31	6.9099683	4.184914	30	7.13	4.81	–0.05	[–0.55; 0.45]
Carey (2006)	84	5.5220312	4.116353	79	6.20	4.00	–0.17	[–0.47; 0.14]
Collins (2002)	47	6.0463139	3.943795	47	6.94	4.75	–0.20	[–0.61; 0.20]
Cucciare (2013)	82	10.0930587	19.102105	68	12.00	20.00	–0.10	[–0.42; 0.22]
Kypri (2004)	40	1.2695500	1.470736	40	2.08	2.05	–0.45	[–0.89; –0.01]
McNally (2003)	24	3.3118764	3.081781	29	4.17	3.15	–0.27	[–0.81; 0.27]
Michael (2006)	47	3.3494574	3.249407	44	4.20	5.30	–0.19	[–0.61; 0.22]
Murphy (2001)	30	2.1778264	1.110103	24	2.45	1.25	–0.23	[–0.77; 0.31]
Neal (2004)	31	1.3298915	1.219690	30	1.50	1.40	–0.13	[–0.63; 0.37]
Voogt (2013)	456	–0.0560054	2.795866	451	0.00	2.79	–0.02	[–0.15; 0.11]
Werch (2000)	266	1.4232671	1.929889	255	1.30	1.60	0.07	[–0.10; 0.24]
White (2008)	111	–0.1472335	0.671476	119	–0.10	0.61	–0.07	[–0.33; 0.19]
Fixed effect model	**1278**			**1246**			**–0.07**	**[–0.15; 0.01]**
Random effects model							**–0.07**	**[–0.15; 0.01]**

Heterogeneity: I–squared=0%, tau–squared=0, p=0.6188

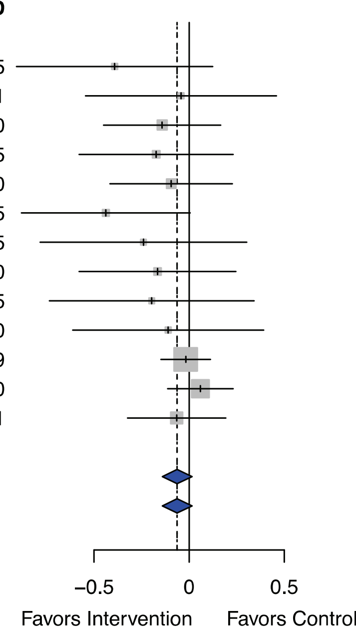


Forest Plot of Frequnecy of Heavy Drinking at 0–3 Months, Full Adjustment of Effect Sizes

	Experimental			Control				
Study	**Total**	**Mean**	**SD**	**Total**	**Mean**	**SD**	**SMD**	**95%–CI**
Borsari (2000)	29	2.836869	1.4325727	30	3.37	1.25	–0.39	[–0.91; 0.12]
Borsari (2005)	31	6.934952	4.1822503	30	7.13	4.81	–0.04	[–0.54; 0.46]
Carey (2006)	84	5.620707	4.1123789	79	6.20	4.00	–0.14	[–0.45; 0.17]
Collins (2002)	47	6.176387	3.9401660	47	6.94	4.75	–0.17	[–0.58; 0.23]
Cucciare (2013)	82	10.142400	19.1000642	68	12.00	20.00	–0.09	[–0.42; 0.23]
Kypri (2004)	40	1.290520	1.4706486	40	2.08	2.05	–0.44	[–0.88; 0.01]
McNally (2003)	24	3.409313	3.0814641	29	4.17	3.15	–0.24	[–0.78; 0.30]
Michael (2006)	47	3.468570	3.2475571	44	4.20	5.30	–0.17	[–0.58; 0.25]
Murphy (2001)	30	2.215942	1.1082732	24	2.45	1.25	–0.20	[–0.73; 0.34]
Neal (2004)	31	1.353714	1.2186289	30	1.50	1.40	–0.11	[–0.61; 0.39]
Voogt (2013)	456	–0.049966	2.7954149	451	0.00	2.79	–0.02	[–0.15; 0.11]
Werch (2000)	266	1.405326	1.9279478	255	1.30	1.60	0.06	[–0.11; 0.23]
White (2008)	111	–0.142140	0.6713584	119	–0.10	0.61	–0.07	[–0.32; 0.19]
Fixed effect model	**1278**			**1246**			**–0.06**	**[–0.14; 0.02]**
Random effects model							**–0.06**	**[–0.14; 0.02]**

Heterogeneity: I–squared=0%, tau–squared=0, p=0.7676

Standardised mean difference



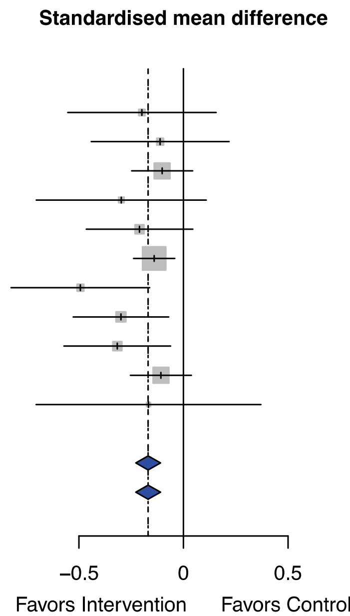
Medium Follow-up (4-16 months)

<u>Frequency of Consumption</u>

Forest Plot of Frequnecy of Alcohol Consumption at 4−16 Months,
No Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%−CI
Carey (2006)	64	4.100	2.50	59	4.60	2.50	−0.20	[−0.55; 0.16]
Cucciare (2013)	75	17.000	10.90	67	18.20	10.50	−0.11	[−0.44; 0.22]
Hansen (2012)	365	−1.400	13.72	358	0.00	13.72	−0.10	[−0.25; 0.04]
Kypri (2004)	47	3.170	1.63	47	3.83	2.65	−0.30	[−0.70; 0.11]
Kypri (2008)	113	3.870	2.72	126	4.45	2.78	−0.21	[−0.46; 0.04]
Kypri (2009)	767	−0.094	0.67	811	0.00	0.67	−0.14	[−0.24; −0.04]
Lewis (2007b)	67	1.860	1.15	78	2.43	1.15	−0.49	[−0.82; −0.16]
Marlatt (1998)	143	2.300	1.00	156	2.60	1.00	−0.30	[−0.53; −0.07]
Martens (2013)	112	2.270	1.41	128	2.74	1.54	−0.32	[−0.57; −0.06]
Moreira (2012)	349	6.040	6.10	369	6.73	6.66	−0.11	[−0.25; 0.04]
Murphy (2001)	30	3.170	1.21	24	3.37	1.14	−0.17	[−0.70; 0.37]
Fixed effect model	**2132**			**2223**			**−0.17**	**[−0.23; −0.11]**
Random effects model							**−0.17**	**[−0.23; −0.11]**

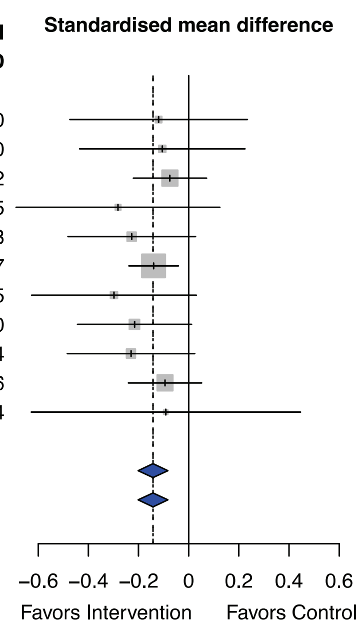
Heterogeneity: I−squared=0%, tau−squared=0, p=0.5666



Forest Plot of Frequnecy of Alcohol Consumption at 4−16 Months,
Moderate Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%−CI
Carey (2006)	64	4.291832	2.598287	59	4.60	2.50	−0.12	[−0.47; 0.23]
Cucciare (2013)	75	17.055835	11.056836	67	18.20	10.50	−0.11	[−0.44; 0.22]
Hansen (2012)	365	−1.031992	13.813741	358	0.00	13.72	−0.07	[−0.22; 0.07]
Kypri (2004)	47	3.200709	1.662930	47	3.83	2.65	−0.28	[−0.69; 0.12]
Kypri (2008)	113	3.819565	2.741671	126	4.45	2.78	−0.23	[−0.48; 0.03]
Kypri (2009)	767	−0.094000	0.670000	811	0.00	0.67	−0.14	[−0.24; −0.04]
Lewis (2007b)	67	2.078689	1.198804	78	2.43	1.15	−0.30	[−0.63; 0.03]
Marlatt (1998)	143	2.379968	1.031535	156	2.60	1.00	−0.22	[−0.44; 0.01]
Martens (2013)	112	2.395284	1.436094	128	2.74	1.54	−0.23	[−0.48; 0.02]
Moreira (2012)	349	6.124737	6.118382	369	6.73	6.66	−0.09	[−0.24; 0.05]
Murphy (2001)	30	3.256594	1.281921	24	3.37	1.14	−0.09	[−0.63; 0.45]
Fixed effect model	**2132**			**2223**			**−0.14**	**[−0.20; −0.08]**
Random effects model							**−0.14**	**[−0.20; −0.08]**

Heterogeneity: I−squared=0%, tau−squared=0, p=0.9499

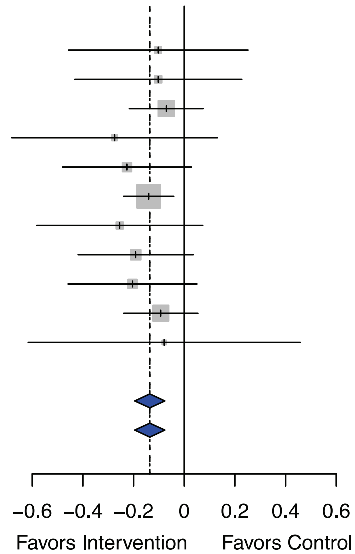


Forest Plot of Frequnecy of Alcohol Consumption at 4−16 Months,
Full Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%−CI
Carey (2006)	64	4.336685	2.594220	59	4.60	2.50	−0.10	[−0.46; 0.25]
Cucciare (2013)	75	17.085440	11.053762	67	18.20	10.50	−0.10	[−0.43; 0.23]
Hansen (2012)	365	−0.975240	13.807717	358	0.00	13.72	−0.07	[−0.22; 0.08]
Kypri (2004)	47	3.216992	1.662466	47	3.83	2.65	−0.27	[−0.68; 0.13]
Kypri (2008)	113	3.823600	2.741426	126	4.45	2.78	−0.23	[−0.48; 0.03]
Kypri (2009)	767	−0.094000	0.670000	811	0.00	0.67	−0.14	[−0.24; −0.04]
Lewis (2007b)	67	2.129821	1.197800	78	2.43	1.15	−0.25	[−0.58; 0.07]
Marlatt (1998)	143	2.404952	1.030796	156	2.60	1.00	−0.19	[−0.42; 0.04]
Martens (2013)	112	2.434425	1.435641	128	2.74	1.54	−0.20	[−0.46; 0.05]
Moreira (2012)	349	6.136600	6.117704	369	6.73	6.66	−0.09	[−0.24; 0.05]
Murphy (2001)	30	3.272476	1.277380	24	3.37	1.14	−0.08	[−0.62; 0.46]
Fixed effect model	**2132**			**2223**			**−0.14**	**[−0.20; −0.08]**
Random effects model							**−0.14**	**[−0.20; −0.08]**

Heterogeneity: I−squared=0%, tau−squared=0, p=0.9772

Standardised mean difference



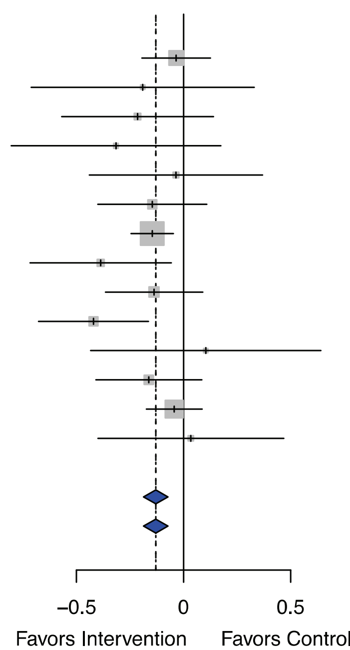
Average Consumption

Forest Plot of Average Consumption at 4−16 Months,
No Adjustment of Effect Sizes

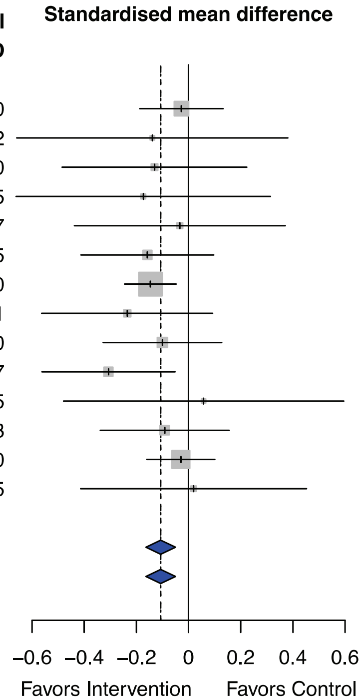
Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%−CI
Bewick (2013)	281	16.500	18.40	321	17.10	16.50	−0.03	[−0.19; 0.13]
Borsari (2005)	29	18.690	9.75	28	21.04	14.22	−0.19	[−0.71; 0.33]
Carey (2006)	64	12.800	9.90	59	15.00	10.50	−0.21	[−0.57; 0.14]
Collins (2002)	33	1.330	0.31	32	1.42	0.25	−0.32	[−0.80; 0.17]
Kypri (2004)	47	8.040	4.75	47	8.23	5.87	−0.04	[−0.44; 0.37]
Kypri (2008)	113	8.280	5.06	126	9.02	5.05	−0.15	[−0.40; 0.11]
Kypri (2009)	767	−0.117	0.80	811	0.00	0.80	−0.15	[−0.25; −0.05]
Lewis (2007b)	67	8.410	6.71	78	11.02	6.71	−0.39	[−0.72; −0.06]
Marlatt (1998)	143	2.400	1.50	156	2.60	1.40	−0.14	[−0.36; 0.09]
Martens (2013)	112	9.570	8.78	128	13.74	10.77	−0.42	[−0.68; −0.16]
Murphy (2001)	30	16.630	9.29	24	15.72	7.75	0.10	[−0.43; 0.64]
Neighbors (2004)	126	8.530	8.76	126	10.01	9.43	−0.16	[−0.41; 0.09]
Voogt (2013)	456	21.500	20.70	451	22.40	20.50	−0.04	[−0.17; 0.09]
Walters (2007)	39	3.170	6.11	43	2.98	4.95	0.03	[−0.40; 0.47]
Fixed effect model	**2307**			**2430**			**−0.13**	**[−0.19; −0.07]**
Random effects model							**−0.13**	**[−0.19; −0.07]**

Heterogeneity: I−squared=0%, tau−squared=0, p=0.4625

Standardised mean difference



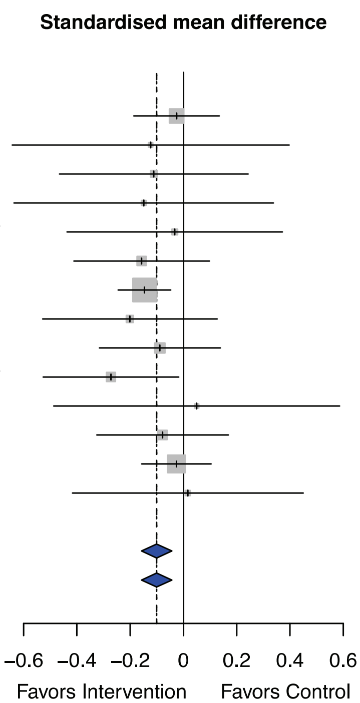
Forest Plot of Average Consumption at 4−16 Months, Moderate Adjustment of Effect Sizes



Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%−CI
Bewick (2013)	281	16.619259	18.716206	321	17.10	16.50	−0.03	[−0.19; 0.13]
Borsari (2005)	29	19.316418	9.978142	28	21.04	14.22	−0.14	[−0.66; 0.38]
Carey (2006)	64	13.644062	10.187349	59	15.00	10.50	−0.13	[−0.48; 0.22]
Collins (2002)	33	1.364530	0.368608	32	1.42	0.25	−0.17	[−0.66; 0.31]
Kypri (2004)	47	8.048841	4.826660	47	8.23	5.87	−0.03	[−0.44; 0.37]
Kypri (2008)	113	8.215652	5.109283	126	9.02	5.05	−0.16	[−0.41; 0.10]
Kypri (2009)	767	−0.117000	0.800000	811	0.00	0.80	−0.15	[−0.25; −0.05]
Lewis (2007b)	67	9.411365	6.909251	78	11.02	6.71	−0.24	[−0.56; 0.09]
Marlatt (1998)	143	2.453312	1.526267	156	2.60	1.40	−0.10	[−0.33; 0.13]
Martens (2013)	112	10.681560	8.884048	128	13.74	10.77	−0.31	[−0.56; −0.05]
Murphy (2001)	30	16.235996	9.565561	24	15.72	7.75	0.06	[−0.48; 0.59]
Neighbors (2004)	126	9.170798	8.895276	126	10.01	9.43	−0.09	[−0.34; 0.16]
Voogt (2013)	456	21.792713	20.791613	451	22.40	20.50	−0.03	[−0.16; 0.10]
Walters (2007)	39	3.087735	6.166686	43	2.98	4.95	0.02	[−0.41; 0.45]
Fixed effect model	**2307**			**2430**			**−0.11**	**[−0.16; −0.05]**
Random effects model							**−0.11**	**[−0.16; −0.05]**

Heterogeneity: I−squared=0%, tau−squared=0, p=0.9044

Forest Plot of Average Consumption at 4−16 Months, Full Adjustment of Effect Sizes



Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%−CI
Bewick (2013)	281	16.648603	18.7011428	321	17.10	16.50	−0.03	[−0.19; 0.13]
Borsari (2005)	29	19.512124	9.9724057	28	21.04	14.22	−0.12	[−0.64; 0.40]
Carey (2006)	64	13.841415	10.1763138	59	15.00	10.50	−0.11	[−0.47; 0.24]
Collins (2002)	33	1.372603	0.3661021	32	1.42	0.25	−0.15	[−0.64; 0.34]
Kypri (2004)	47	8.053528	4.8250393	47	8.23	5.87	−0.03	[−0.44; 0.37]
Kypri (2008)	113	8.220800	5.1087008	126	9.02	5.05	−0.16	[−0.41; 0.10]
Kypri (2009)	767	−0.117000	0.8000000	811	0.00	0.80	−0.15	[−0.25; −0.05]
Lewis (2007b)	67	9.645497	6.9042535	78	11.02	6.71	−0.20	[−0.53; 0.13]
Marlatt (1998)	143	2.469968	1.5254466	156	2.60	1.40	−0.09	[−0.32; 0.14]
Martens (2013)	112	11.028833	8.8837199	128	13.74	10.77	−0.27	[−0.53; −0.02]
Murphy (2001)	30	16.163734	9.5440797	24	15.72	7.75	0.05	[−0.49; 0.59]
Neighbors (2004)	126	9.288322	8.8871780	126	10.01	9.43	−0.08	[−0.33; 0.17]
Voogt (2013)	456	21.858200	20.7848531	451	22.40	20.50	−0.03	[−0.16; 0.10]
Walters (2007)	39	3.072648	6.1624732	43	2.98	4.95	0.02	[−0.42; 0.45]
Fixed effect model	**2307**			**2430**			**−0.10**	**[−0.16; −0.04]**
Random effects model							**−0.10**	**[−0.16; −0.04]**

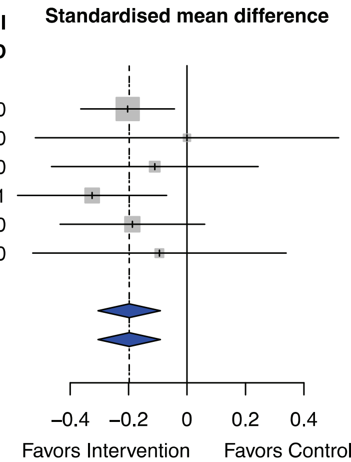
Heterogeneity: I−squared=0%, tau−squared=0, p=0.9468

Peak BAC

Forest Plot of Peak BAC at 4−16 Months, No Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%−CI
Bewick (2013)	281	8.440	4.870	321	9.500	5.490	−0.20	[−0.36; −0.04]
Borsari (2005)	29	0.170	0.120	28	0.170	0.140	0.00	[−0.52; 0.52]
Carey (2006)	64	0.160	0.080	59	0.170	0.100	−0.11	[−0.46; 0.24]
Martens (2013)	112	0.111	0.089	128	0.144	0.111	−0.32	[−0.58; −0.07]
Neighbors (2004)	126	7.660	5.160	126	8.650	5.390	−0.19	[−0.43; 0.06]
Walters (2007)	39	0.050	0.110	43	0.060	0.100	−0.09	[−0.53; 0.34]
Fixed effect model	**651**			**705**			**−0.20**	**[−0.30; −0.09]**
Random effects model							**−0.20**	**[−0.30; −0.09]**

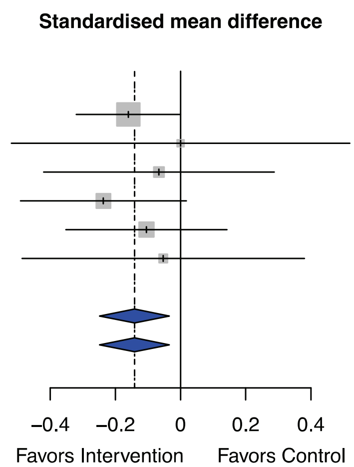
Heterogeneity: I−squared=0%, tau−squared=0, p=0.8533



Forest Plot of Peak BAC at 4−16 Months, Moderate Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%−CI
Bewick (2013)	281	8.65069002	5.05948016	321	9.500	5.490	−0.16	[−0.32; 0.00]
Borsari (2005)	29	0.17000000	0.12172204	28	0.170	0.140	0.00	[−0.52; 0.52]
Carey (2006)	64	0.16383665	0.08411434	59	0.170	0.100	−0.07	[−0.42; 0.29]
Martens (2013)	112	0.11979651	0.09019484	128	0.144	0.111	−0.24	[−0.49; 0.02]
Neighbors (2004)	126	8.08864185	5.29079191	126	8.650	5.390	−0.10	[−0.35; 0.14]
Walters (2007)	39	0.05432972	0.11101128	43	0.060	0.100	−0.05	[−0.49; 0.38]
Fixed effect model	**651**			**705**			**−0.14**	**[−0.25; −0.03]**
Random effects model							**−0.14**	**[−0.25; −0.03]**

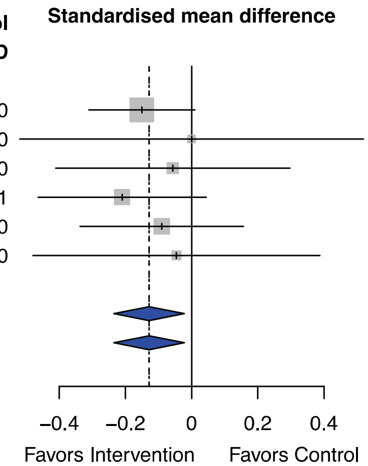
Heterogeneity: I−squared=0%, tau−squared=0, p=0.9355



Forest Plot of Peak BAC at 4−16 Months,
Full Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%−CI
Bewick (2013)	281	8.7025323	5.05187809	321	9.500	5.490	−0.15	[−0.31; 0.01]
Borsari (2005)	29	0.1700000	0.12165193	28	0.170	0.140	0.00	[−0.52; 0.52]
Carey (2006)	64	0.1647337	0.08391916	59	0.170	0.100	−0.06	[−0.41; 0.30]
Martens (2013)	112	0.1225447	0.09018022	128	0.144	0.111	−0.21	[−0.46; 0.04]
Neighbors (2004)	126	8.1672562	5.28304505	126	8.650	5.390	−0.09	[−0.34; 0.16]
Walters (2007)	39	0.0551238	0.11094336	43	0.060	0.100	−0.05	[−0.48; 0.39]
Fixed effect model	**651**			**705**			**−0.13**	**[−0.24; −0.02]**
Random effects model							**−0.13**	**[−0.24; −0.02]**

Heterogeneity: I−squared=0%, tau−squared=0, p=0.9548

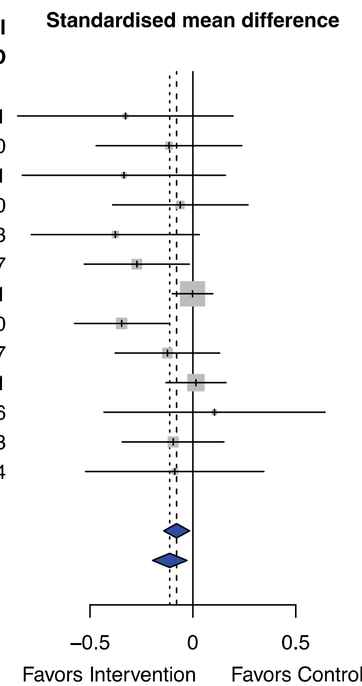


Alcohol Related Problems

Forest Plot of Alcohol Related Problems at 4−16 Months,
No Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%−CI
Borsari (2005)	29	5.000	5.09	28	6.71	5.21	−0.33	[−0.85; 0.20]
Carey (2006)	64	4.700	5.20	59	5.30	5.10	−0.12	[−0.47; 0.24]
Collins (2002)	32	6.800	9.53	32	9.77	7.91	−0.34	[−0.83; 0.16]
Cucciare (2013)	75	5.900	10.20	67	6.50	9.30	−0.06	[−0.39; 0.27]
Kypri (2004)	47	2.620	1.91	47	3.45	2.43	−0.38	[−0.78; 0.03]
Kypri (2008)	113	2.570	1.99	126	3.17	2.37	−0.27	[−0.53; −0.02]
Kypri (2009)	767	−0.001	0.71	811	0.00	0.71	0.00	[−0.10; 0.10]
Marlatt (1998)	143	4.000	4.00	156	5.50	4.60	−0.35	[−0.57; −0.12]
Martens (2013)	112	7.760	9.42	128	8.91	9.17	−0.12	[−0.38; 0.13]
Moreira (2012)	349	18.050	2.01	369	18.02	2.01	0.01	[−0.13; 0.16]
Murphy (2001)	30	6.460	3.51	24	6.07	3.86	0.10	[−0.43; 0.64]
Neighbors (2004)	126	5.740	7.18	126	6.45	7.63	−0.10	[−0.34; 0.15]
Walters (2007)	39	1.510	2.30	43	1.72	2.44	−0.09	[−0.52; 0.35]
Fixed effect model	**1926**			**2016**			**−0.08**	**[−0.14; −0.02]**
Random effects model							**−0.11**	**[−0.20; −0.03]**

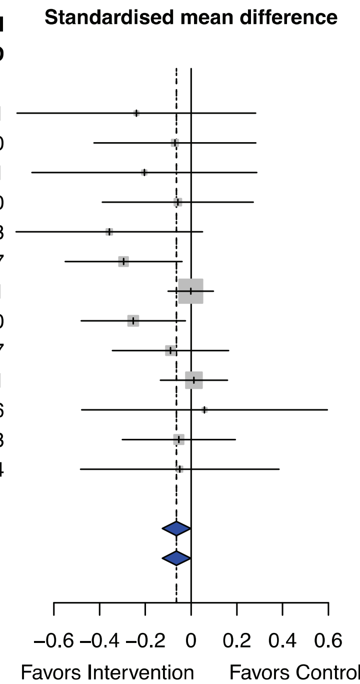
Heterogeneity: I−squared=24.9%, tau−squared=0.0052, p=0.1922



Forest Plot of Alcohol Related Problems at 4–16 Months, Moderate Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%–CI
Borsari (2005)	29	5.455819	5.142306	28	6.71	5.21	−0.24	[−0.76; 0.28]
Carey (2006)	64	4.930199	5.299336	59	5.30	5.10	−0.07	[−0.42; 0.28]
Collins (2002)	32	7.939484	9.692576	32	9.77	7.91	−0.20	[−0.70; 0.29]
Cucciare (2013)	75	5.927918	10.257183	67	6.50	9.30	−0.06	[−0.39; 0.27]
Kypri (2004)	47	2.658619	1.933435	47	3.45	2.43	−0.36	[−0.77; 0.05]
Kypri (2008)	113	2.517826	2.003882	126	3.17	2.37	−0.29	[−0.55; −0.04]
Kypri (2009)	767	−0.001000	0.710000	811	0.00	0.71	0.00	[−0.10; 0.10]
Marlatt (1998)	143	4.399842	4.042266	156	5.50	4.60	−0.25	[−0.48; −0.02]
Martens (2013)	112	8.066545	9.498157	128	8.91	9.17	−0.09	[−0.34; 0.16]
Moreira (2012)	349	18.046316	2.243016	369	18.02	2.01	0.01	[−0.13; 0.16]
Murphy (2001)	30	6.291141	3.618113	24	6.07	3.86	0.06	[−0.48; 0.60]
Neighbors (2004)	126	6.047410	7.270554	126	6.45	7.63	−0.05	[−0.30; 0.19]
Walters (2007)	39	1.600924	2.325212	43	1.72	2.44	−0.05	[−0.48; 0.38]
Fixed effect model	**1926**			**2016**			**−0.06**	**[−0.13; 0.00]**
Random effects model							**−0.06**	**[−0.13; 0.00]**

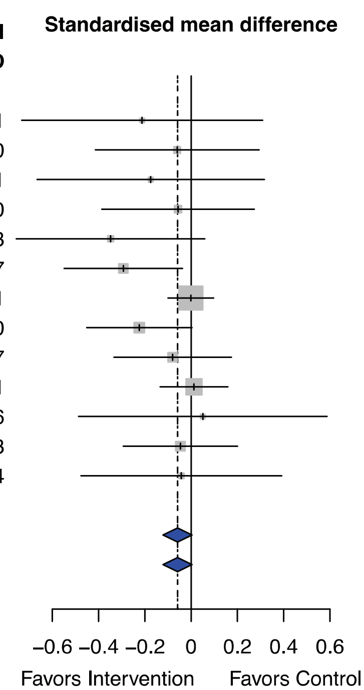
Heterogeneity: I–squared=0%, tau–squared=0, p=0.4984



Forest Plot of Alcohol Related Problems at 4–16 Months, Full Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%–CI
Borsari (2005)	29	5.598226	5.141584	28	6.71	5.21	−0.21	[−0.73; 0.31]
Carey (2006)	64	4.984022	5.294696	59	5.30	5.10	−0.06	[−0.41; 0.29]
Collins (2002)	32	8.205910	9.687647	32	9.77	7.91	−0.17	[−0.67; 0.32]
Cucciare (2013)	75	5.942720	10.255991	67	6.50	9.30	−0.06	[−0.39; 0.27]
Kypri (2004)	47	2.679096	1.933151	47	3.45	2.43	−0.35	[−0.76; 0.06]
Kypri (2008)	113	2.522000	2.003733	126	3.17	2.37	−0.29	[−0.55; −0.04]
Kypri (2009)	767	−0.001000	0.710000	811	0.00	0.71	0.00	[−0.10; 0.10]
Marlatt (1998)	143	4.524760	4.041814	156	5.50	4.60	−0.22	[−0.45; 0.00]
Martens (2013)	112	8.162316	9.495720	128	8.91	9.17	−0.08	[−0.33; 0.17]
Moreira (2012)	349	18.045800	2.234412	369	18.02	2.01	0.01	[−0.13; 0.16]
Murphy (2001)	30	6.260172	3.609611	24	6.07	3.86	0.05	[−0.49; 0.59]
Neighbors (2004)	126	6.103790	7.264649	126	6.45	7.63	−0.05	[−0.29; 0.20]
Walters (2007)	39	1.617600	2.323544	43	1.72	2.44	−0.04	[−0.48; 0.39]
Fixed effect model	**1926**			**2016**			**−0.06**	**[−0.12; 0.00]**
Random effects model							**−0.06**	**[−0.12; 0.00]**

Heterogeneity: I–squared=0%, tau–squared=0, p=0.6034

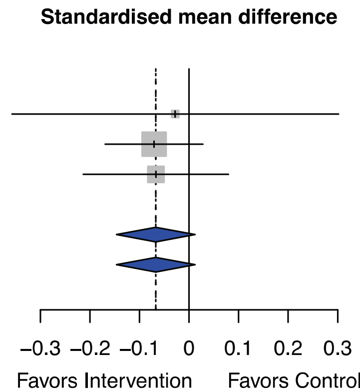


Average Consumption per Occasion

Forest Plot of Average Consumption per Occasion at 4−16 Months,
No Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%−CI
Cucciare (2013)	75	3.900	3.70	67	4.00	3.40	−0.03	[−0.36; 0.30]
Kypri (2009)	767	−0.041	0.58	811	0.00	0.58	−0.07	[−0.17; 0.03]
Moreira (2012)	349	4.320	2.09	369	4.46	2.09	−0.07	[−0.21; 0.08]
Fixed effect model	**1191**			**1247**			**−0.07**	**[−0.15; 0.01]**
Random effects model							**−0.07**	**[−0.15; 0.01]**

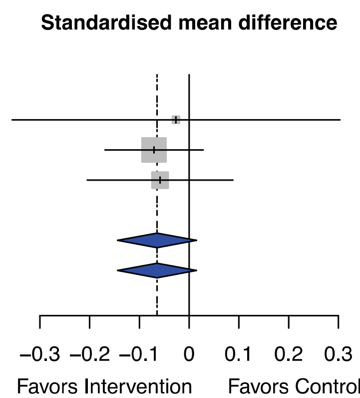
Heterogeneity: I−squared=0%, tau−squared=0, p=0.9708



Forest Plot of Average Consumption per Occasion at 4−16 Months,
Moderate Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%−CI
Cucciare (2013)	75	3.904653	3.732722	67	4.00	3.40	−0.03	[−0.36; 0.30]
Kypri (2009)	767	−0.041000	0.580000	811	0.00	0.58	−0.07	[−0.17; 0.03]
Moreira (2012)	349	4.337193	2.106602	369	4.46	2.09	−0.06	[−0.20; 0.09]
Fixed effect model	**1191**			**1247**			**−0.06**	**[−0.14; 0.01]**
Random effects model							**−0.06**	**[−0.14; 0.01]**

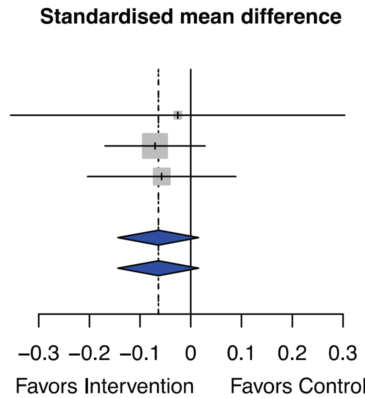
Heterogeneity: I−squared=0%, tau−squared=0, p=0.9644



Forest Plot of Average Consumption per Occasion at 4–16 Months, Full Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%–CI
Cucciare (2013)	75	3.90712	3.732021	67	4.00	3.40	–0.03	[–0.36; 0.30]
Kypri (2009)	767	–0.04100	0.580000	811	0.00	0.58	–0.07	[–0.17; 0.03]
Moreira (2012)	349	4.33960	2.105974	369	4.46	2.09	–0.06	[–0.20; 0.09]
Fixed effect model	**1191**			**1247**			**–0.06**	**[–0.14; 0.02]**
Random effects model							**–0.06**	**[–0.14; 0.02]**

Heterogeneity: I–squared=0%, tau–squared=0, p=0.9622

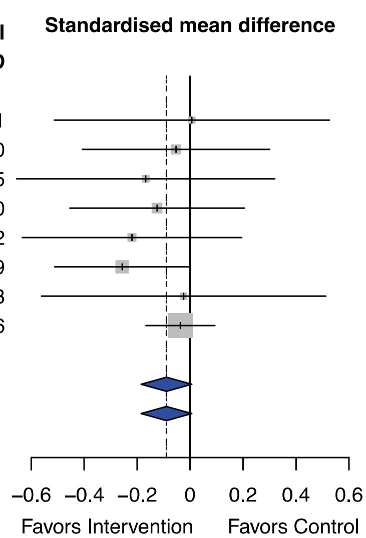


Frequency of Heavy Drinking

Forest Plot of Frequency of Heavy Drinking at 4–16 Months, No Adjustment of Effect Sizes

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%–CI
Borsari (2005)	29	6.100	4.07	28	6.07	4.71	0.01	[–0.51; 0.53]
Carey (2006)	64	4.900	3.50	59	5.10	4.00	–0.05	[–0.41; 0.30]
Collins (2002)	33	6.360	4.55	32	7.22	5.55	–0.17	[–0.65; 0.32]
Cucciare (2013)	75	11.000	23.00	67	14.00	25.00	–0.12	[–0.45; 0.21]
Kypri (2004)	45	1.510	1.27	45	1.91	2.22	–0.22	[–0.63; 0.20]
Kypri (2008)	113	1.190	1.18	126	1.60	1.89	–0.26	[–0.51; 0.00]
Murphy (2001)	30	1.870	1.11	24	1.90	1.33	–0.02	[–0.56; 0.51]
Voogt (2013)	456	–0.105	2.86	451	0.00	2.86	–0.04	[–0.17; 0.09]
Fixed effect model	**845**			**832**			**–0.09**	**[–0.19; 0.01]**
Random effects model							**–0.09**	**[–0.19; 0.01]**

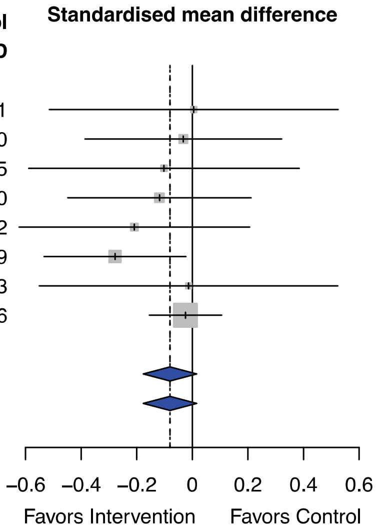
Heterogeneity: I–squared=0%, tau–squared=0, p=0.8827



Forest Plot of Frequency of Heavy Drinking at 4–16 Months,
Moderate Adjustment of Effect Sizes

	Experimental			Control				
Study	**Total**	**Mean**	**SD**	**Total**	**Mean**	**SD**	**SMD**	**95%–CI**
Borsari (2005)	29	6.0920032	4.132907	28	6.07	4.71	0.00	[–0.51; 0.52]
Carey (2006)	64	4.9767329	3.605816	59	5.10	4.00	–0.03	[–0.39; 0.32]
Collins (2002)	33	6.6899517	4.693835	32	7.22	5.55	–0.10	[–0.59; 0.38]
Cucciare (2013)	75	11.1395881	23.119023	67	14.00	25.00	–0.12	[–0.45; 0.21]
Kypri (2004)	45	1.5286117	1.283003	45	1.91	2.22	–0.21	[–0.62; 0.21]
Kypri (2008)	113	1.1543478	1.186199	126	1.60	1.89	–0.28	[–0.53; –0.02]
Murphy (2001)	30	1.8829891	1.141510	24	1.90	1.33	–0.01	[–0.55; 0.52]
Voogt (2013)	456	–0.0708502	2.866014	451	0.00	2.86	–0.02	[–0.15; 0.11]
Fixed effect model	**845**			**832**			**–0.08**	**[–0.18; 0.02]**
Random effects model							**–0.08**	**[–0.18; 0.02]**

Heterogeneity: I–squared=0%, tau–squared=0, p=0.8166



Forest Plot of Frequency of Heavy Drinking at 4–16 Months,
Full Adjustment of Effect Sizes

	Experimental			Control				
Study	**Total**	**Mean**	**SD**	**Total**	**Mean**	**SD**	**SMD**	**95%–CI**
Borsari (2005)	29	6.089505	4.130313	28	6.07	4.71	0.00	[–0.51; 0.52]
Carey (2006)	64	4.994674	3.600351	59	5.10	4.00	–0.03	[–0.38; 0.33]
Collins (2002)	33	6.767099	4.687886	32	7.22	5.55	–0.09	[–0.57; 0.40]
Cucciare (2013)	75	11.213600	23.116687	67	14.00	25.00	–0.12	[–0.45; 0.21]
Kypri (2004)	45	1.538480	1.282813	45	1.91	2.22	–0.20	[–0.62; 0.21]
Kypri (2008)	113	1.157200	1.186135	126	1.60	1.89	–0.28	[–0.53; –0.02]
Murphy (2001)	30	1.885371	1.139300	24	1.90	1.33	–0.01	[–0.55; 0.53]
Voogt (2013)	456	–0.063210	2.865552	451	0.00	2.86	–0.02	[–0.15; 0.11]
Fixed effect model	**845**			**832**			**–0.08**	**[–0.17; 0.02]**
Random effects model							**–0.08**	**[–0.17; 0.02]**

Heterogeneity: I–squared=0%, tau–squared=0, p=0.8171