Placebo group improvement in trials of pharmacotherapies for alcohol use disorders: A multivariate meta-analysis examining change over time

AC Del Re, PhD, Natalya Maisel, PhD, Janet Blodgett, MA, Paula Wilbourne, PhD, and John Finney, PhD

VA Palo Alto Health Care System, USA
Stanford University Medical School, USA

[PRE-PEER REVIEW (PRE-PRINT) VERSION OF ARTICLE]

Corresponding author:

AC Del Re, PhD
Health Services Research Fellow
HSR&D Center for Health Care Evaluation
Stanford University Medical School
VA Palo Alto Health Care System (152MPD)
795 Willow Rd
Menlo Park, CA 94025
acdelre@stanford.edu
650-493-5000 x2-25842
Fax: 650- 617-2736

Note: Supplemental material is included with this manuscript. This file includes references for all studies included in the meta-analysis.

Acknowledgments

Preparation of this manuscript was supported by U.S. National Institute on Alcohol Abuse and Alcoholism Grant No. AA008689 and the U.S. Department of Veterans Affairs, Office of Research and Development, Health Services Research and Development Service and Substance Use Disorder Quality Enhancement Research Initiative. The views expressed are those of the authors and do not necessarily represent the views of the National Institute on Alcohol Abuse and Alcoholism, the Department of Veterans Affairs, or any other U.S. Government entity. We thank Helen Pettinati for her helpful comments on an earlier version of this manuscript.

Abstract

Objective: Placebo group improvement in pharmacotherapy trials has been increasing over time across several pharmacological treatment areas. However, it is unknown to what degree increasing improvement has occurred in pharmacotherapy trials for alcohol use disorders or what factors may account for placebo group improvement. This meta-analysis of 47 alcohol pharmacotherapy trials evaluated (1) the magnitude of placebo group improvement, (2) the extent to which placebo group improvement has been increasing over time, and (3) several potential moderators that may account for variation in placebo group improvement.

Method: Random-effects univariate and multivariate analyses were conducted that examined the magnitude of placebo group improvement in the 47 studies and several potential moderators of improvement: (a) publication year, (b) country in which the study was conducted, (c) outcome data source/type, (d) number of placebo administrations, (e) overall severity of study participants, and (f) additional psychosocial treatment.

Results: Substantial placebo group improvement was found overall, and improvement was larger in more recent studies. Greater improvement was found on moderately subjective outcomes, with more frequent administrations of the placebo, and in studies with greater participant severity of illness. However, even after controlling for these moderators, placebo group improvement remained significant, as did placebo group improvement over time.

Conclusion: Similar to previous pharmacotherapy placebo research, substantial pre- to post-test placebo group improvement has occurred in alcohol pharmacotherapy trials, an effect that has been increasing over time. Increasing placebo group improvement over time persisted even after controlling for several potential moderators of this relationship.

Placebo group improvement in trials of pharmacotherapies for alcohol use disorders: A multivariate meta-analysis examining change over time

Establishing the efficacy of pharmacotherapies typically involves a series of double-blind, controlled trials in which participants are randomized to a medication or a placebo (control) group. The treatment group receives the active medication while the placebo group is administered an inactive substance that otherwise mimics the medication and its administration. Including a placebo control group accounts for the effects of time (i.e., natural reduction or regression in symptoms over time), regression to the mean from measurement error, and the meaning connected with taking a (potential) medication.

Previous research has shown that the meaning associated with treatment and the treatment context, such as a doctor's attire (e.g., white coat) and administration of an intervention purported to heal (whether or not it contains the active ingredient), can influence outcomes [1]. Such study features can impact participants' hope, beliefs and expectations of getting better [2,3]. These effects are subsumed under such terms as the "placebo effect," "meaning response," or "placebo response." In the context of this paper, which focuses on pharmacotherapy trials for alcohol use disorders, such effects are included in the improvement made between baseline and end-of-treatment in the placebo control groups in clinical trials. Substantial placebo group improvement, and increasing placebo group improvement over time, may be responsible for the relatively small positive effects found in meta-analyses for pharmacological treatments for alcohol use disorders, such as naltrexone and acamprosate [4–7], and the decline in such effects over time (e.g., Feinn & Kranzler[8]).

Controversy Regarding the Placebo Response

The most interesting component of placebo group improvement, the placebo response, has a controversial history, with some researchers (e.g., Hróbjartsson & Gotzsche[9]) claiming it is mostly irrelevant in affecting outcomes, and others (e.g., Wampold et al.[3]) deeming it to have a "powerful" influence on outcomes across several types of treatments. Specifically, Hróbjartsson and Gotzsche[9] conducted a meta-analysis of trials of pharmacological and psychosocial interventions for several medical and psychiatric conditions that, in addition to a treatment group, contained both placebo and no-treatment groups. They found an overall mean placebo effect (outcome difference between placebo and no-treatment groups) close to zero, indicating a virtually nonexistent placebo response. The only exception was in pain studies in which a small but significant placebo effect was found.

In contrast, Wampold et al. [3] reanalyzed the studies reviewed by Hróbjartsson and Gotzsche. They found that when the research design was appropriate (i.e., the treatment was indistinguishable from placebo and participants were aware of receiving some form of treatment) and the disorder was amenable to placebos (e.g., insomnia, pain, depression, and other conditions that have psychological components that are expected to be affected by patient expectations versus such conditions as a bacterial infection or anemia that are unlikely to be affected by patient expectations), the placebo response was robust [10–13].

Is Improvement in Placebo Groups Increasing Over Time?

The placebo response and broader placebo group improvement have remained topics of considerable interest in more recent research and in the popular media [14,15]. Interestingly, researchers are finding that placebo group improvement has been increasing over time across several pharmacological treatment areas, including medications for schizophrenia (Kemp et al.[16]) and depression [17]. We examine whether year of publication is a moderator of the magnitude of

placebo group improvement in alcohol medication trials, with the hypothesis that placebo group improvement is larger in more recent studies.

Increasing Placebo Group Improvement and Moderators of Placebo Group Improvement

Although the mechanisms underlying placebo group improvement and increasing placebo group improvement over time are not well understood, several explanations have been proposed.

Marketing exposure. Perhaps the most interesting predictor (moderator) was offered by one author in the popular press who speculated that more intensive marketing and advertisements for pharmacotherapies may lead to an increasing placebo response over time [14]. Consumers in the United States have been exposed to progressively more media advertisements for prescription medications since the FDA amended its direct-to-consumer advertising in 1997. The messages claim positive outcomes for medications in treating a wide range of conditions. Potentially, these advertisements could induce an expectancy response and influence medication trial outcomes by generating an inflated placebo response.

This marketing exposure hypothesis is plausible because meaning-making is heavily influenced by cultural norms, ideas, symbols, and expectations [1,3]. Meaning is the driving force behind hope and expectation, and these psychological factors are central concepts in the theory of the placebo response [2]. Therefore, greater exposure over time to messages about the efficacy of medications should increase both individual consumers' beliefs and expectations, as well as cultural norms and ideas about the efficacy of medications [3]. One approach to testing this hypothesis is to examine the location (country) where a medication trial is conducted. One would expect the placebo response to be stronger in a country, such as the USA, with direct-to-consumer advertising compared to other countries where direct-to-consumer advertising is illegal.

In fact, a stronger placebo response has been reported for pharmacotherapy trial sites within the USA versus sites outside the USA in at least one review [16].

Source and type of outcome data. Kemp et al.[16] hypothesized that the increasing trend over time in placebo group improvement in schizophrenia medication trials may be related to the increasing use of less reliable, subjective reports of outcomes. Self-reported outcomes, especially those involving ratings of internal states rather than behavior, should be influenced more by individuals' hope, beliefs and expectations more than are self-reports of more objective behaviors and outcomes assessed by more objective means than self-report. However, the empirical evidence is mixed. For example, Hróbjartsson and Gotzsche [9] found that self-reported (subjective) outcome measures had larger improvement effect sizes than objective measures (e.g., those completed by an external rater) and that improvement on subjective pain outcomes differed significantly from zero. However, Wampold et al. [3] found no differences by source/type of outcome data.

Number of placebo administrations. The number of placebo administrations has been found in previous research to be positively correlated with improvement on outcome variables [18,19]. More placebo administrations may increase patient expectations for improvement, which then results in better outcomes. If the number of placebo administrations in alcohol trials has been increasing over time, it could also account for hypothesized trends of increasing placebo group improvement.

Severity of illness. Patient symptom severity has been proposed to play a role in placebo group improvement, as some authors note greater placebo group improvement may simply reflect greater regression to the mean in studies with greater baseline patient severity [11,20]. Thus, a trend toward more placebo group improvement over time may be related to changes in trial inclusion

criteria that have increasingly admitted patients with greater severity [16,17]. Rief et al.[17] note that severity cut-off scores for anti-depressant study inclusion have increased over time and, for investigators to obtain a sufficient number of participants, they may initially overestimate patient severity for selection into the trial. Given regression to the mean, this exaggeration of patient baseline symptom severity will inevitably result in lower scores on symptom measures upon reassessment. However, in the review by Reif et al.[17], even after controlling for diagnosis and severity of depression, the year of study publication remained a significant predictor of outcome. We will examine whether severity accounts for placebo group improvement and increasing improvement over time in our sample of alcohol pharmacotherapy trials.

Additional treatment. Finally, provision of additional treatment, such as psychosocial treatment, in conjunction with a placebo administration in medication trials should increase placebo group improvement above that of placebo-only administration [11]. If the use of psychosocial treatment in pharmacotherapy trials has been increasing over time, it also could account for the trend of increasing placebo group improvement.

Current Study

Although the placebo response and placebo group improvement have been studied in medication trials for several disorders, it is unknown to what degree placebo group improvement occurs in pharmacotherapy trials for alcohol use disorders or if there has been a similar increasing trend in placebo group improvement over time. If placebo group improvement is increasing over time and if medication and placebo effects are not independent, but are overlapping, then the detection of medication effects may prove more difficult over time. Assuming a "cap" on the amount of improvement that can be achieved with use of alcohol medications as currently formulated, there would be less room for medication effects to be identified if placebo group

improvements are increasing over time. That is, if treatment and placebo effects are not producing mutually exclusive effects (i.e., are not additive; see Wampold et al.[3] for detailed explanation), then increasing placebo group improvements over time would "subtract from" the improvements that otherwise could have been produced by the medication.

The purpose of the current meta-analysis is to examine placebo group improvement in pharmacological treatments for alcohol use disorders. It is hypothesized that placebo control groups in such trials exhibit significant improvement from pre- to post-testing. We also determine the magnitude of the (presumed to be negative) relationship between placebo group improvement and medication-placebo between-group effect sizes on drinking-related outcomes. A second hypothesis is that the magnitude of placebo group improvement has been increasing over time. In addition to publication year, we examine several other potential moderators of the magnitude of placebo group improvement, including trial location, number of placebo administrations, participant baseline severity, source of outcome data, and the provision of additional treatment in conjunction with placebo. To gauge whether these moderator variables account for any increase in the placebo response over time, their relationships to year of publication and the extent to which they account for any relationship of publication year to placebo response are examined.

Methods

Literature Search

Randomized controlled trials (RCTs) of pharmacotherapy for alcohol dependence were identified through multiple searches of PubMed and PsycINFO conducted at different points over the past decade, reflecting the intermittent availability of funds and resources. We used search terms for various medications (e.g., “naltrexone”), terms for alcohol problems and dependence

(e.g., “alcohol*,” “problem drinking”) and terms for randomized controlled trials (e.g., “randomized controlled,” “clinical trial”). Study inclusion criteria were (a) a focus on treating alcohol misuse or an alcohol use disorder; (b) participants 18 years of age or older; (c) publication between 1970 and 2009; (d) a report in the English language; and (e) random assignment of at least five participants each to medication and placebo groups.

For example, in one search on PubMed, we used the term “alcoholism,” terms for pharmacotherapy (“pharmacotherapy OR drug treatment OR drug therapy”) and terms for randomized controlled trial (“randomized controlled OR randomized clinical OR randomized trial* OR controlled trial OR clinical trial OR placebo OR double-blind method”), with limits for “Human subjects” and reports in the “English language” (we did not have the resources to translate reports in other languages). This search yielded 1,602 potential research reports. Based on examination of abstracts and full text versions of these reports, 1,184 were identified as not relevant (e.g., qualitative studies, reviews). Of the remaining articles, 215 were rejected based on our inclusion criteria (e.g., open-label trial), 138 articles met the inclusion criteria, and 65 articles were duplicate publications for studies already in the dataset (e.g., secondary analysis).

In addition to the database searches, we consulted the reference sections from the reports of the included studies and from previously published reviews of this literature. Across these search strategies, a total of 161 studies met our inclusion criteria. For the present analysis, trials were excluded if the report contained no data on baseline scores on the outcome variable, or no significance tests on within-group pre- to post-test change, or no narrative information on pre- to post-test change (k = 114). The remaining 47 studies constituted the data set for this meta-analysis. Some of the medications examined were naltrexone, acamprosate, sertraline, and topiramate (a list of the 47 trials is available from the first author).

Moderator Coding

Moderator variables in meta-analyses are study characteristics that may explain some or all of the between-study variability in effect sizes (ESs). Several categorical and continuous moderators were coded and are discussed below. Each moderator was double-coded and consensus was reached between the coders and the project coordinator (NM).

- *Publication year (k*=47) was coded as a continuous variable and indicates the year of study publication.
- *Trial location* (*k*= 47) was based on the country location where the trial was conducted (USA=1, Other=0).
- *Outcome data source/type* (*k*= 68) was coded into three categories: (1) highly subjective measure (*k*=17), (2) moderately subjective measure (k=36), or (3) objective measure (*k*=15). Highly subjective outcome measures were self-reports that assessed participants' internal *experiences* (e.g., craving for alcohol). Moderately subjective outcomes were self-reports of alcohol-related *behaviors* (e.g., frequency and quantity of alcohol use). Objective measures were physiological markers of alcohol consumption.
- *Number of placebo administrations per day* (*k*=36) was coded as a continuous variable signifying the total number of placebo administrations that participants received each day (e.g., the number of times during the day that the participant was supposed to take a pill). Eleven studies did not report any data for this variable, limiting the number of studies reporting this moderator to *k*=36. When placebo medications were administered on a time interval other than daily administrations (e.g., weekly), those values were divided appropriately to obtain a daily estimate.

- *Severity of illness* (*k*=46) was coded into three categorical levels: (1) non-clinical problem drinkers (e.g., drinkers with clear alcohol-related problems, but not dependent), (2) clinical problem drinkers (e.g., dependent on alcohol, but not severely dependent), and (3) severely-impaired dependent drinkers. This coding scheme was based on the severity codes for the "mesa grande" review by Miller and Wilbourne[21] in which coders rated severity on a 5-point scale. However, participant severity in all the studies in the current review fell into one of the three categories above. Only one study was of "non-clinical problem drinkers," so it was dropped from all analyses containing this moderator variable. Therefore, the final severity of illness moderator had only two factor levels: Clinical problem drinkers and severely-impaired dependent drinkers.
- *Additional treatment* (*k*=47) was coded as a dichotomous variable: (0) placebo only, (1) placebo plus psychosocial treatment. Coders determined if one or more psychosocial intervention(s) were provided to placebo group (and medication) participants in conjunction with the medication trial (yes/no). The psychosocial interventions included medication management [22], supportive counseling, twelve-step facilitation, and cognitive-behavioral therapy, among others.

Calculation of Effect Sizes

The ES metric for this meta-analysis was Hedges' [23] *g*. To compute *g*, we first calculated Cohen's [24] *d* for placebo control conditions using Equation 1,

$$d = \frac{\bar{X}_{post} - \bar{X}_{pre}}{sd_{pre}}, \quad (1)$$

where $\bar{X}_{post}$ and $\bar{X}_{pre}$ are the mean scores on the dependent variable at post- and pre-treatment, respectively, and sd_{pre} is the standard deviation pre-treatment. Scores on "negative outcomes,"

such as days of binge drinking, were reversed, so in all cases a positive ES indicated improvement in functioning of the placebo group. The variance of *d* is given by:

$$V_d = \left(\frac{1}{n} + \frac{d^2}{2n}\right) \cdot 2(1 - r), \quad (2)$$

where *n* is the group size and *r* is the correlation between pre- and post-scores. Because *r* generally was not reported, we assumed a pre-post correlation of *r* = .5 for all studies, after conducting sensitivity analyses with a range or correlation values between *r* = .3 to *r* =.7, none of which influenced the results to any substantial degree. We then used the procedure recommended by Hedges [25] to correct both *d* and V_d for small sample bias, giving us *g* and V_g.

Effect sizes for studies not reporting means and standard deviations. The majority of studies we analyzed reported pre- and post-test means and standard deviations (*k* = 41) for the placebo group, which allowed us to compute ESs using the formulae in Equations 1 and 2. For studies not reporting these values, ESs typically were computed from *t* or *F* statistics for pre-post change, using the R statistical software package 'compute.es' [26]. One study (Chick et al.[27]) reported non-significant findings, but no statistics that would allow computation of *g*. We entered *g* = 0 in this case. Finally, four studies reported significant improvement, but no statistic that would allow one to compute an ES. For these studies, we made the conservative assumption that *p*=.05 and computed the corresponding value of *g*.

Dependent effect sizes. For each drinking-related outcome reported in the original source, one baseline to end-of-treatment Hedges' *g* ES was computed. If multiple outcomes were reported, the ESs within studies were aggregated to obtain one ES for each independent sample. Accounting for the within-study correlation among ESs is a recommended[28,29] but rarely utilized practice in meta-analytic research, due to lack of reporting of relevant data in the primary studies, software limitations and/or computational complexity. However, if it is not done, the

consequence is meta-analytic results with such "inferior statistical properties" [30(p20)] as biased point estimates and overestimated variance of the summary ES.

Although availability of all between-measure correlations within each study would be ideal, an imputation procedure for missing values is recommended when correlations are not available [30]. Such imputation has been utilized in several meta-analyses [31–33] and was also used in this meta-analysis, given that the correlations among the outcome measures in the primary studies were not reported. To determine the imputation value, a sensitivity analyses used several values ranging from *r*=.3 to *r*=.7. The imputed ES values did not significantly differ from one another and resulted in inconsequential differences in the findings of omnibus and moderator analyses. Therefore, an imputed value of *r*=.5 was used, as has been the case in other meta-analyses, although the correlations in those studies were among within-study, between-group ESs, rather than within-group, pre-post ESs (for details on this procedure see[29,34,35]). The R statistical software package 'MAd' [36] was used to conduct this aggregation and to carry out several additional analyses.

When comparing across different *outcome data sources/types*, we used a multivariate approach. Whereas all other moderators (e.g., publication year) had only one value per study, some studies used more than one source/type of outcome data. For example, Guardia et al[37] and Balldin et al.[38] reported two and three sources/types of outcome data, respectively. For studies having outcomes assessed by more than one data source/type, treating them as providing independent data is not justifiable. Therefore, a multivariate procedure that accounts for correlated variables was used. This procedure is discussed in more depth in the Statistical Analysis section below.

Statistical Analyses

A random effects restricted maximum-likelihood estimator was generated in both univariate[39] and multivariate analyses [15], which assumes that the studies in a meta-analysis were randomly sampled from a population of studies. All analyses were conducted using R statistical software packages - univariate methods with the 'MAd' [36] package, multivariate meta-analytic methods with the 'mvmeta' [40] package, and meta-analytic diagnostics (tests for outliers and influential cases) with the 'metafor' [39] package.

Univariate models. The first random-effects univariate analysis involved an unconditional (i.e., omnibus) model (not conditioned on study level variables, i.e., moderators), as follows:

$$\theta_j = \mu + v_j^* \quad (3)$$

where μ is the average true pre-post change and $v_j^* = v_j + \tau^2$, where the variance of the within-study errors v_j is known and the between-study errors τ^2 are unknown and estimated from the studies included in the analysis. Homogeneity was tested with the *Q*-statistic and indexed as a percentage of variance in study findings due to true differences with the I^2-statistic [25,41]. *Q* has an approximate χ^2 distribution with *k* - 1 degrees of freedom, where *k* is the number of studies aggregated. *Q*-values of above the critical value result in rejection of the null hypothesis of homogeneity.

A conditional model was then used to examine the effects of each moderator variable individually, prior to implementing multipredictor models. For example, publication year moderator (Year) was entered in a meta-regression model as follows:

$$\theta_j = \gamma_0 + \gamma_1(Year) + v_j^* \quad (4)$$

where γ_0 is the expected placebo group improvement for a study when the moderator is zero, centered at the grand mean or centered in another way. For these analyses, Year was centered based on the publication year for first available study in the dataset which was 1985, so γ_1 is the

expected difference in effect size per a unit change of the moderator (i.e., a change from 1985 to 1986 and so on). If a moderator variable accounts for some of the variability in study effects, the fixed effect γ_1 will be significantly different than zero (p-value < .05) and the variance, , will be reduced. Remaining heterogeneity in these moderator analysis models also was assessed.

Similarly, when several potentially relevant moderators were included in a univariate multi-predictor meta-regression, the formula yields

$$\theta_j = \gamma_0 + \gamma_1(Year) + \gamma_2(Trial\ location) + \gamma_3(Data\ source) + \gamma_4(Administrations) + \gamma_5(Severity) + \gamma_6(Additional) + \nu_j^* \quad (5)$$

in the case in which Year, Trial location, Outcome data source (*Data source*), Number of placebo administrations (*Administrations*), Severity, and Additional treatment (*Additional*) are entered simultaneously. In this equation example, γ_0 is the expected effect for a trial conducted outside of the USA using subjective outcomes with no additional treatment and all continuous moderators (i.e., Year, Administrations, and Severity moderators) grand mean centered. In this model, γ_1 ...γ_3 are the expected differences in effect size per unit change each of the moderators, while holding the effect of each of the other moderators (γ_j) constant. If, for example, γ_1 remains statistically significant in this model, it can be inferred that the Year is a robust moderator in the sense that it is not fully "explained" by the other moderators.

Multivariate models. The outcome data source/type moderator could not be aggregated further without losing relevant information and resulting in a dependent dataset (i.e., more than one ES per study). Univariate methods are statistically inappropriate for dependent datasets as they assume independence of within-study ESs, which realistically are correlated and not independent. Therefore, we examined outcome data source/type (highly subjective vs. moderately subjective vs. objective outcome measures) as a moderator using multivariate meta-analytic

methods [40] which yield estimates that account for covariance between within-study ESs. Therefore, they do not violate the statistical assumption of independence. Computational details on the multivariate meta-analysis approach can be found in Jackson, Riley, and White[42].

Results

Prior to conducting the main omnibus and moderator analyses, preliminary diagnostics were run to identify outlier studies. We first examined the distribution of the ESs. One study's effect size[43] deviated substantially from the others in the distribution of placebo group ESs (indicating that two percent of this sample was outliers). Whereas the remaining 46 ESs ranged from g = 0.0 to g = 2.27, this study's g = 3.04. However, based on sampling theory, five percent of a distribution is expected to be outliers (i.e., beyond the 95% tails of the distribution). This expectation, coupled with this study's ES not being an extreme outlier, led us to include the study in all analyses.

In addition, for the omnibus and final multipredictor models, we examined standardized residuals for each study as recommended by Hedges and Olkin[25] (pp. 258-259) to identify ESs that are not well accounted for by a given model (i.e., studies that may have exerted undue influence on estimates of model coefficients). In most analyses, none of these standardized residuals (scaled as z scores) exceeded an absolute value of 2, suggesting no outliers or highly influential studies. In the few cases where the residuals exceeded an absolute value of 2, we tested and compared the models including and excluding the extreme residuals. In all cases, the significance of moderators did not change and model coefficient values were minimally impacted.

Unconditional model. The overall effect of the unconditional (omnibus) model analysis (k=47) was g_+ = .899 (95% CI = .704, 1.09), indicating there was a "large" (based on Cohen's

interpretive guidelines; 1988) and significant placebo group improvement from pre- to post-testing. In other words, the average end-of-treatment outcome score for the placebo group was nearly a full standard deviation higher than the average baseline score. In addition, the correlation between placebo group improvement and mean differences between medication and placebo (Hedges' *g* between medication and placebo group means at post-test) was negative, as expected (r = -.201). This correlation is relatively small, but it indicates that larger placebo group improvements are modestly related to smaller detectable post-test mean differences between medication and placebo groups.

Although placebo group improvement was significant in the omnibus model, there was a large degree of heterogeneity in effect sizes (Q = 541.62, p < .0001; I^2 = 91%; see "Intercept only" row in Table 3 for details). Therefore, we conducted moderator analyses to determine if one or more study characteristics could explain the variability in placebo group improvement. Prior to running single and multi-predictor meta-regressions, we examined the correlations between moderators, using Pearson's *r* for continuous variables, a Point-biserial *r* for relationships between a dichotomous and a continuous variable, and Phi *r* for relationships between two dichotomous variables (see Table 3). None of the other moderators was significantly related to year of publication. Only the relationship between number of administrations and trial location was statistically significant, indicating that studies conducted outside of the USA typically had fewer placebo administrations.

Single-moderator analyses. Results of the single-predictor moderator analyses are presented in Table 4 (univariate models) and Table 5 (multivariate models for outcome data source/type). As seen in Table 4, the effect of year of publication was significant and in the expected direction (larger placebo group improvement in more recently published trials). In

contrast, location of the trial, frequency of placebo administration, severity of participants' illness, and provision of psychosocial treatment in addition to placebo did not moderate (explain) pre-to-post changes in the placebo group in single-moderator analyses. As seen in Table 5, improvement given each outcome data source/type was significantly larger than zero (highly subjective $y_{0=}$ 0.770, moderately subjective $y_{0=}$ 1.312. objective $y_{0=}$ 0.599). However, when comparing outcomes using a Wald test based on a standard normal distribution (not shown in Table 5), the significant effect of outcome data source/type indicated that improvement on moderately subjective outcomes was significantly larger than that on subjective outcomes (p=0.002), but improvement on subjective outcomes was not significantly different than that on objective outcomes (p=0.115).

Multiple-moderator analysis. All moderators (except the multivariate outcome data source/type) were entered simultaneously as predictors of effect size in a multiple moderator meta-regression. As shown in Table 6, the multiple moderator model explained significant heterogeneity in effect sizes (Q(5) = 12.191; p = .032). Several moderators, including year of publication ($y_{1=}$ 0.040, p=.045), number of administrations ($y_{1=}$ 0.304, p=.048), and patient severity ($y_{1=}$ 0.218, p=.039), were significant in the final model, all in the expected direction. However, trial location and additional treatment were not significant in the final model.

The significance of the publication year moderator coefficient indicates that placebo group improvement has been increasing over time, and that this effect cannot be explained by the other moderators in the model. The number of placebo administrations coefficient value indicates, after controlling for the other predictors, placebo group improvement was larger in trials with more placebo administrations. Finally, the patient severity coefficient value indicates

that, controlling for the other moderators, trials with more severe patients exhibited greater placebo group improvement.

Discussion

This random-effects meta-analysis examined placebo group improvement in pharmacological treatment trials for alcohol use disorders. Each of the findings will be discussed as they relate to the literature and theory of the placebo response and placebo group improvements.

Placebo group improvement and increases over time. Similar to previous studies [3,16] and extending to pharmacotherapy trials of alcohol use disorders, this meta-analysis found, as hypothesized, significant placebo group improvement from a baseline assessment to the end of treatment. As would be expected, greater placebo group improvement was associated with smaller medication versus placebo effects at the end of treatment, suggesting that placebo group improvement may be responsible for some, but not a major portion, of the perceived reduction in medication effects over time [8], as the magnitude of this relationship was modest. Further, placebo group improvement in alcohol pharmacotherapy trials has gotten stronger over time (i.e., more recent studies showed more placebo group improvement), and this effect persisted even after controlling for several additional moderators.

Marketing exposure. It was hypothesized that the moderator of escalating marketing exposure (i.e., increasing exposure to advertisements espousing the efficacy of pharmacotherapy), due to the U. S. FDA amendment to its direct-to-consumer advertising in 1997, is responsible for increasing placebo group improvements over time. Although our indicator of media exposure, trial location, showed a trend in the expected direction, the moderator test proved non-significant. Therefore, although the marketing exposure idea is

compelling, our operationalization of it failed to explain why placebo group improvement has been increasing over time. This finding is in contrast to both the findings of Kemp et al.[16] and our hypothesis. We cannot rule out the possibility that a more sensitive indicator of actual marketing exposure in all of the countries in which alcohol medication trials have been conducted would account for increasing placebo group improvement in the trials that were reviewed. Further, we recognize there could be confounds with the trial location moderator. Specifically, direct marketing for alcohol medications in the USA is virtually nonexistent and there may be stigma associated with treating alcohol dependence with another drug (e.g., Naltrexone). Previous research has shown that both patient and provider attitudes about pharmacotherapy for alcohol use disorders are conflicted, particularly about the possibility of "trading one addiction for another" [44(p287)]. This perception, although not tested empirically, could attenuate placebo group changes in American trials, resulting in an overall reduction in detecting ES differences between countries (e.g., USA vs. Other countries).

Outcome data source/type. Outcomes involving self-reports of internal states rather than behavior were expected to be influenced more by individuals' hope, beliefs and expectations than outcomes assessed by more objective means, but this hypothesis was not supported. Although placebo group improvement was significantly greater than zero for outcomes with all sources/types of data, improvement on moderately subjective outcomes (i.e., self-reports of drinking behavior) was significantly larger than that on subjective outcomes (i.e., self-report of the internal state of alcohol craving), but improvement on subjective outcomes was not significantly different than that on objectively assessed outcomes (i.e., physiological markers). This finding is in contrast to at least some findings from previous research and to our hypothesis that highly subjective outcomes would exhibit larger effects than other sources. The difference in

our results relative to previous findings may be accounted for by the specific types of outcome variables examined in alcohol treatment trials. It seems possible that persons in treatment may reduce or eliminate the behavior of alcohol consumption but still have craving for alcohol over an extended period, accounting for why less improvement may have been found on self-reported craving relative to self-reported drinking behavior outcomes. Assuming this is the case, investigators mounting pharmacotherapy trials might consider using additional outcome types and data sources in conjunction with behavioral self-reports of alcohol use. Presumably, outcomes that are less sensitive to the placebo group changes (but still responsive to medication) would yield stronger medication effects.

Number of placebo administrations. More placebo administrations were associated with stronger placebo group improvement after holding all other moderators constant. In fact, each additional administration per day was associated with an improvement ES increase of about one-third of a standard deviation. Research from several decades ago also found a positive relationship between the number of placebo administrations and improvement on outcome variables [18,19]. More placebo administrations may increase patient expectations for improvement, which then results in better outcomes. This finding has implications for design of future pharmacotherapy trials for alcohol use disorders. For example, trial coordinators may want to reduce the impact of placebo group improvement. Reducing this impact might be achieved by administering placebo medications on a less frequent basis, if formulations permit. In contrast, providers may wish to harness the full power of their patients' psychological resources (in conjunction with the therapeutic effects of the medication), by increasing expectancy effects through use of more frequent drug administrations, again as formulations permit.

Severity of illness. A severity of alcohol dependence moderator variable was used to try to account for regression to the mean (RTM) in placebo group improvement. However, the effects associated with this moderator are a combination of both RTM and the actual influence of illness severity on outcomes. Two equally valid interpretations can be made of the significance of this moderator: (1) placebo group improvement is substantially impacted by RTM and (2) participants with greater severity of illness are more responsive to placebos. Although each of these interpretations in isolation is plausible, both of them are probably valid. That is, RTM likely is occurring to some degree *and* participants with greater severity of illness likely respond more favorably to a placebo. Perhaps the latter response is due to a stronger desire for relief from suffering and greater sensitivity to anything perceived as providing relief. This finding is in contrast to the generally held notion that participant's with light/moderately severe AUDs are more responsive to placebo group improvements because, it is thought, they do not really need medication to improve.

Additional treatment. The hypothesis that provision of additional treatment in conjunction with a placebo would increase placebo group improvement above that of placebo-only treatment was not supported. The findings are somewhat surprising, given one might expect greater improvements in placebo groups receiving additional treatment. Although it is not uncommon to have moderator groups of this size, the total number of studies (K =7) in the additional treatment group may have been small enough to slightly inflate this group's ES standard error, resulting in somewhat less sensitivity to detecting group differences.

Suppression effects. Several moderators were *not* significant in the univariate models, but were in multi-moderator analysis. This pattern reflects other moderators acting as “suppressor variables.” A suppressor variable is defined as a "variable which increases the predictive validity

of another variable (or set of variables) by its inclusion into a regression equation" [2] (p. 3). Suppression in this case means that instead of an expected drop in coefficient significance when adding additional moderators, the opposite happened. For example, in the single-predictor model, number of placebo administrations was not statistically significant. However, when adding the other moderators simultaneously in a multi-predictor model, number of administrations became a statistically significant predictor of placebo group improvement. Part of this effect is likely due to reducing error variance in this moderator which then will increase the proportion of variance remaining that can contribute to the prediction of placebo group improvement.

Limitations. A primary limitation of our review is that we only examined placebo group improvements and not the specific placebo response (or placebo effect). Placebo group improvement includes the placebo response, but also several other factors, such as regression to the mean due to measurement error and fluctuations in symptoms over time, with selection into trials typically at a point when symptoms are especially severe. To examine the specific placebo response requires comparing changes in a placebo group to changes in a no-treatment control group. This was not feasible, as only a small proportion of studies included a no-treatment control condition, which would have limited statistical power to detect effects along with limiting the generalizability of findings. Another limitation is that less than a third of the otherwise relevant alcohol pharmacotherapy trials provided sufficient information on placebo group improvement. It is not known to what extent the findings on the subset of studies would generalize to the entire set. Finally, we examined placebo group response at the study level. Factors other than the moderators examined here may account for at least some of the variation in the improvement of placebo group members within studies.

Future research. Future reviews should investigate the placebo response in isolation from placebo group improvement by examining differences between placebo and no-treatment groups. This will be possible when there are a sufficient number of primary studies that include a no-treatment control group along with a medication and placebo group. However, even with a limited number of studies reporting either placebo or no-treatment group findings, future syntheses should consider using network meta-analysis (or mixed-treatment meta-analysis; MTC) to estimate the placebo response. MTC can be used to assess the comparative effectiveness of treatments among similar participant populations (e.g., participants with alcohol use disorders) that have not been compared directly in a trial (typically an RCT). MTC combines data from all comparisons (direct and indirect) among a set of several treatments while accounting for (1) the possibility of bias in the indirect comparison and (2) the degree of agreement between direct and indirect comparisons from the data.

Conclusion

As hypothesized, this meta-analysis found significant placebo group improvement from a baseline assessment to the end of treatment in pharmacotherapy trials of alcohol use disorders. As expected, greater placebo group improvement was associated with smaller medication versus placebo effects at the end of treatment, suggesting that placebo group improvement may be responsible for some of the perceived reduction in medication effects over time (e.g., Feinn & Kranzler[8]). Further, placebo group improvement in alcohol pharmacotherapy trials have become stronger over time (i.e., more recent studies showed greater placebo group improvement) and this effect persisted even after controlling for several other moderators. In addition, more placebo administrations and greater patient severity were associated with more placebo group

improvement. Pharmacotherapy researchers may be able to use these results to more effectively isolate alcohol medication effects.

Acknowledgments

Preparation of this manuscript was supported by U.S. National Institute on Alcohol Abuse and Alcoholism Grant No. AA008689 and the U.S. Department of Veterans Affairs, Office of Research and Development, Health Services Research and Development Service and Substance Use Disorder Quality Enhancement Research Initiative. The views expressed are those of the authors and do not necessarily represent the views of the National Institute on Alcohol Abuse and Alcoholism, the Department of Veterans Affairs, or any other U.S. Government entity. We thank ___and Helen Pettinati for her helpful comments on an earlier version of this manuscript.

References

1. Moerman DE, Jonas WB. Deconstructing the placebo effect and finding the meaning response. *Annals of Internal Medicine*. 2002;136(6):471–476.

2. Tracey I. Getting the pain you expect: mechanisms of placebo, nocebo and reappraisal effects in humans. *Nature medicine*. 2010;16(11):1277–1283.

3. Wampold BE, Minami T, Tierney SC, Baskin TW, Bhati KS. The placebo is powerful: estimating placebo effects in medicine and psychotherapy from randomized clinical trials. *Journal of Clinical Psychology*. 2005;61(7):835–854.

4. Carmen B, Angeles M, Ana M, María AJ. Efficacy and safety of naltrexone and acamprosate in the treatment of alcohol dependence: a systematic review. *Addiction*. 2004;99(7):811–828.

5. Garbutt JC, West SL, Carey TS, Lohr KN, Crews FT. Pharmacological treatment of alcohol dependence. *JAMA: The Journal of the American Medical Association*. 1999;281(14):1318.

6. Kranzler HR. Evidence-Based Treatments for Alcohol Dependence. *JAMA: the journal of the American Medical Association*. 2006;295(17):2075.

7. Kranzler HR, Kirk J. Efficacy of Naltrexone and Acamprosate for Alcoholism Treatment: A Meta-Analysis. *Alcoholism: Clinical and Experimental Research*. 2001;25(9):1335–1341.

8. Feinn R, Kranzler HR. Does Effect Size in Naltrexone Trials for Alcohol Dependence Differ for Single-Site vs. Multi-Center Studies? *Alcoholism: Clinical and Experimental Research*. 2005;29(6):983–988.

9. Hróbjartsson A, Gotzsche PC. Is the placebo powerless? *New England Journal of Medicine*. 2001;344(21):1594–1602.

10. Hróbjartsson A, Gotzsche PC. Placebo interventions for all clinical conditions. *Cochrane Database Syst Rev*. 2010;1.

11. Kienle GS, Kiene H. The powerful placebo effect: fact or fiction? *Journal of Clinical Epidemiology*. 1997;50(12):1311–1318.

12. Wampold BE, Imel ZE, Minami T. The placebo effect:“Relatively large” and “robust” enough to survive another assault. *Journal of clinical psychology*. 2007;63(4):401–403.

13. Wampold BE, Imel ZE, Minami T. The story of placebo effects in medicine: evidence in context. *Journal of clinical psychology*. 2007;63(4):379–390.

14. Silberman S. Placebos are getting more effective: drugmakers are desperate to know why. *Wired Magazine*. 2009;17:1–8.

15. Vance E. Seeking to Illuminate the Mysterious Placebo Effect. *The New York Times*. 2010. Available at: http://www.nytimes.com/2010/06/22/health/22prof.html. Accessed November 29, 2011.

16. Kemp AS, Schooler NR, Kalali AH, et al. What is causing the reduced drug-placebo difference in recent schizophrenia clinical trials and what can be done about it? *Schizophrenia bulletin*. 2010;36(3):504–509.

17. Rief W, Nestoriuc Y, Weiss S, et al. Meta-analysis of the placebo response in antidepressant trials. *Journal of affective disorders*. 2009;118(1-3):1–8.

18. Blackwell B, Bloomfield SS, Buncher CR. Demonstration to medical students of placebo responses and non-drug factors. *The Lancet*. 1972;299(7763):1279–1282.

19. Lasagna L, Mosteller F, von Felsinger JM, Beecher HK. A study of the placebo response. *The American journal of medicine*. 1954;16(6):770–779.

20. Finney JW. Regression to the mean in substance use disorder treatment research. *Addiction*. 2008;103(1):42–52.

21. Miller WR, Wilbourne PL. Mesa Grande: a methodological analysis of clinical trials of treatments for alcohol use disorders. *Addiction*. 2002;97(3):265–277.

22. Donovan D, Ernst D, Miller W, Pettinati H, Weiss R. Medical Management Treatment Manual: A Clinical Research Guide for Medically Trained Clinicians Providing Pharmacotherapy as Part of the Treatment for Alcohol Dependence. Combine Monograph Series, Volume 2. 2007.

23. Cooper HM, Hedges LV, Valentine JC. *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation Publications; 2009.

24. Cohen J. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum; 1988.

25. Hedges LV, Olkin I, Statistiker M, Olkin I, Olkin I. *Statistical methods for meta-analysis*. Academic Press Orlando, FL; 1985.

Del Re, A. C. (2010). *compute.es: Compute effect sizes*. R package version 0.2 (http://CRAN.R-project.org/package=compute.es)

27. Chick J, Anton R, Checinski K, et al. A multicentre, randomized, double-blind, placebo-controlled trial of naltrexone in the treatment of alcohol dependence or abuse. *Alcohol Alcohol*. 2000;35(6):587–93.

28. Hunter JE, Schmidt FL. *Methods of meta-analysis: Correcting error and bias in research findings*. Sage Publications, Inc; 2004.

29. Gleser LJ, Olkin I. Stochastically dependent effect sizes. *The handbook of research synthesis*. 1994:339–355.

30. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2009;172(4):789–811.

31. Farber BA, Doolin EM. Positive regard. *Psychotherapy*. 2011;48(1):58.

32. Flückiger C, Del Re A, Wampold BE, Symonds D, Horvath AO. How central is the alliance in psychotherapy? A multilevel longitudinal meta-analysis. 2011.

33. Wampold BE, Budge S, Laska KM, et al. Evidence-based treatments for depression and anxiety versus treatment-as-usual: A meta-analysis of direct comparisons. *Clinical Psychology Review*. 2011.

34. Horvath AO, Del Re A, Flückiger C, Symonds D. Alliance in individual psychotherapy. *Psychotherapy*. 2011;48(1):9.

35. Wampold BE, Mondin GW, Moody M, et al. A meta-analysis of outcome studies comparing bona fide psychotherapies: Empiricially," all must have prizes." *Psychological bulletin*. 1997;122(3):203.

Del Re, A. C., & Hoyt, W. T. (2010). MAd: Meta-analysis with mean differences. R package version 0.8 (http://CRAN.R-project.org/package=MAd)

37. Guardia J, Caso C, Arias F, et al. A double-blind, placebo-controlled study of naltrexone in the treatment of alcohol-dependence disorder: results from a multicenter clinical trial. *Alcohol Clin Exp Res*. 2002;26(9):1381–7.

38. Balldin J, Berglund M, Borg S, et al. A 6-month controlled naltrexone study: combined effect with cognitive behavioral therapy in outpatient treatment of alcohol dependence. *Alcohol Clin Exp Res*. 2003;27(7):1142–9.

39. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*. 2010;36(3).

40. Gasparrini A. *mvmeta: multivariate meta-analysis and meta-regression (2-01)*.

41. Raudenbush SW, Bryk AS. *Hierarchical linear models: Applications and data analysis methods*. Sage Publications, Inc; 2002.

42. Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. *Statistics in Medicine*. 2011.

43. Addolorato G, Leggio L, Ferrulli A, et al. Effectiveness and safety of baclofen for maintenance of alcohol abstinence in alcohol-dependent patients with liver cirrhosis: randomised, double-blind controlled study. *Lancet*. 2007;370(9603):1915–22.

44. Mark TL, Kranzler HR, Poole VH, et al. Barriers to the use of medications to treat alcoholism. *The American Journal on Addictions*. 2003;12(4):281–294.

45. MacKinnon DP, Krull JL, Lockwood CM. Equivalence of the mediation, confounding and suppression effect. *Prevention Science*. 2000;1(4):173–181.

Tables

Table 1
Categorical moderator frequencies

Moderator	*k*
Trial location	47
USA	29
Other assignment	18
Outcome Data Source	68
Highly subjective	17
Moder. subjective	36
Objective	15
Additional treatment	47
No	7
Yes	40
Severity of Illness	47
Non-clinical problem drinkers	1
Clinical problem drinkers	21
Severely-impaired dep. drinkers	25

Note. All studies were coded by two coders; *k* = number of studies in each category; Due to Severity of illness factor-level "Non-clinical problem drinkers" having only 1 study, it was dropped from all analyses.

Table 2.
Continuous Moderator Characteristics: Descriptives and Interrater Reliability

Moderator	*K*	*Min*	*Max*	*M*	*SD*
Publication Year	47	1985	2009	n/a	n/a
Number Admin.	36	.033	3	1.446	0.810

Note. *k* = number of studies providing information on each characteristic (total *k* = 47).

Table 3.

Correlations Between Independent Moderators

	Publication Year	Trial location	Number Admin.	Severity
Publication Year				
Trial location	-0.22			
Number Admin.	-0.20	0.37*		
Severity of Illness	-0.17	0.23	0.13	
Additional Treat.	0.16	-0.16	0.17	0.31

Note. * = *p*-value < 0.05, Pearson's correlation computed for bivariate relationship between continuous moderators, Point-biserial correlation computed for bivariate relationship between continuous and categorical moderators, and Phi correlation computed for bivariate relationship between two categorical moderators. Trial location is coded dichotomously where 1=USA and 0=Other. Additional Treat. is coded dichotomously where 0=No and 1=Yes. Severity of Illness is coded dichotomously where 0 =No and 1=Yes. Severity of Illness is coded dichotomously where 0= Clinical problem drinkers and 1= Severely-impaired dep. drinkers.

Table 4.

Univariate Omnibus and Single-Moderator Analyses

	k	y_0	y_1	95% *CI*	*z*	*P*
Omnibus	47	0.900		[0.709, 1.088]	9.300	<.001
Publication Year	47	0.242	0.036	[0.001, 0.071]	2.030	.043*
Trial location	47	0.830	0.114	[-0.278, 0.506]	0.570	.569
Number Admin.	36	0.820	0.197	[-0.078, 0.473]	1.404	.160
Severity of Illness	47	0.789	0.224	[-0.164, 0.611]	1.131	.258
Additional. Treat.	47	0.600	0.360	[-0.159,0.879]	1.360	.174

Note. Publication year centered at the first publication in this analysis (1985). Number Admin is grand mean centered. Trial location is coded dichotomously where 1=USA and 0=Other. Additional Treat. is coded dichotomously where 0=No and 1=Yes. Severity of Illness is coded dichotomously where 0= Clinical problem drinkers and 1= Severely-impaired dep. drinkers. Univariate analyses used a mixed model (studies random, levels of moderator variables fixed); * = *p*-value < 0.05, *k* = number of studies, y_0 = intercept; y_1 = slope; *z* = *z* statistic for y_1 (except Intercept). The categorical moderator is tabulated separately.

Table 5.

Multivariate Analysis

	k	y_0	95% *CI*	$z(y_0)$	*p*
Outcome data source	68				<.002*
Highly subjective	17	0.770	[0.082, 1.457]	2.194	.0282*
Moder. subjective	36	1.312	[0.814, 1.811]	5.162	< .0001*

Objective	15	0.599	[0.015, 1.184]	2.010	.044*

Note. Multivariate analyses used a random effects model; * = *p*-value < 0.05, *k* = number of studies, y_0 = intercept for each outcome level; $z(y_0)$ = *z* statistic for y_0. Significant continuous moderators are tabulated separately.

Table 6
Multiple-Moderator Analyses (k=35)

	y	95% CI	*p*
Intercept	0.628	[0.014, 1.228]	.045*
Publication year	0.040	[0.001, 0.082]	.029*
Trial location	0.142	[-0.310, 0.594]	.540
Number Admin.	0.304	[0.003, 0.605]	.048*
Severity of Illness	0.218	[0.011, 0.425]	.039*
Additional Treat.	0.140	[-0.498, 0.773]	.672

Note. Overall tests of model significance were $Q(5) =$, $p = .032$; * $p < .05$. Publication year, and Num. Admin are all grand mean centered. Trial location is coded dichotomously where 1=USA and 0=Other. Additional Treat. is coded dichotomously where 0=No and 1=Yes. Severity of illness was contrast coded which sum to 0. Univariate analyses used a mixed model (studies random, levels of moderator variables fixed).