



Published in final edited form as:

Alcohol Clin Exp Res. 2013 June ; 37(6): 1064–1068. doi:10.1111/acer.12067.

The declining efficacy of naltrexone pharmacotherapy for alcohol use disorders over time: A multivariate meta-analysis

AC Del Re, PhD, Natalya Maisel, PhD, Janet Blodgett, MA, and John Finney, PhD
VA Palo Alto Health Care System, USA, Stanford University Medical School, USA

Abstract

Background—Oral naltrexone is an FDA-approved medication for treating alcohol use disorders. Although its efficacy has been supported in multiple clinical trials, an earlier review found that its effect sizes on relapse to heavy drinking and, to a lesser extent, percent days drinking, were smaller in more recent trials and in multi-center than in single-site studies. We examined whether these findings held when studies from 2004 to 2009 were taken into account, and whether single-site versus multi-center trials, the use of placebo run-in periods, and placebo group improvement accounted for variation in naltrexone effects and decreasing effects over time.

Methods—A multivariate meta-analysis of naltrexone pharmacotherapy trials for alcohol use disorders was conducted. All analyses simultaneously modeled effect sizes on outcomes of percent days abstinent and relapse to heavy drinking. Potential moderators of medication effects that were examined included publication year, multi-center design (vs. single site), placebo run-in period, and placebo group improvement.

Results—Statistically significant between-group differences on percent days abstinent (the inverse of percent days drinking) and relapse to heavy drinking favored naltrexone over placebo. Year of publication was a significant moderator for both outcomes, with more recent trials having smaller effect sizes. Neither multi- versus single-site study, the interaction between multi- versus single-site study and year of publication, nor placebo run-in period was a significant moderator of naltrexone effects. Although placebo group improvement was modestly associated with smaller between-group naltrexone versus placebo effect sizes, only 21 studies provided usable information on placebo group improvement. Within those studies, there was no relationship between naltrexone ESs and time, so placebo group improvement was not examined as a moderator of that relationship.

Conclusion—Naltrexone effect sizes have attenuated over time. Moderators that explain why effects have been decreasing remain to be determined.

Keywords

Naltrexone; Multivariate Meta-analysis; Alcohol Treatment Research

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, 650-493-5000 x2-25842, Fax: 650- 617-2736, acdelre@stanford.edu.

Natalya C. Maisel, Ph.D., Research Health Science Specialist, Center for Health Care Evaluation, VA Palo Alto Health Care System, 795 Willow Road (152-MPD), Menlo Park, CA 94025, Phone: (650) 493-5000 (press 2 for Menlo Park, ext. 26966), Fax: (650) 617-2736

Janet Blodgett, MSc, Research Associate, Center for Health Care Evaluation, VA Palo Alto Health Care System, 795 Willow Road (152-MPD), Menlo Park, CA 94025, Janet.Blodgett@va.gov Tel: (650) 617-2736 x27292 Fax: (650) 617-2736

John W. Finney, PhD., Research Health Science Specialist, HSR&D Center for Health Care Evaluation, VA Palo Alto Health Care System (152MPD), 795 Willow Road, Menlo Park, CA 94025, (650) 493-5000 Ext. 2-22848 Office, (650) 617-2736 Fax

Oral naltrexone, an opioid antagonist medication, is currently one of only four FDA-approved medications for treating alcohol use disorders (Harris et al., 2010). Although numerous studies and meta-analyses have supported its efficacy, a meta-analysis by Feinn and Kranzler (2005) found that, compared to earlier studies, more recent trials had smaller effects for relapse to heavy drinking and percent drinking days, although the trend for the latter outcome was not quite significant. In addition, multi-center trials (compared to single-site studies) had smaller medication effects on relapse to heavy drinking.

Previous reviews of pharmacotherapy trials for depression and schizophrenia have come to similar conclusions (Kemp et al., 2010; Rief et al., 2009). That is, effect sizes in those trials both attenuated over time and were smaller in multi-center studies. Several factors have been pointed to that may account for these findings. The attenuation of effects in more recent pharmacotherapy trials may be attributed to (a) regression to the mean as a function of publication bias for earlier studies (Finney, 2008), (b) confounding of more recent studies with multi-center designs (Feinn and Kranzler, 2005), and (c) increasing placebo group improvement over time (Del Re et al., in submission). Multi-center studies are thought to have smaller effects due to variability in (a) study execution (e.g., varying fidelity to treatment protocol, recruitment procedures) among centers (Feinn and Kranzler, 2005) and (b) other contextual factors across sites (e.g., variation in participant and provider characteristics) which then attenuate the aggregated effect size (ES).

The current review extends Feinn and Kranzler's (2005) meta-analysis by using multivariate meta-analytic methods to examine the efficacy of naltrexone (vs. placebo) in a larger sample of single-site and multi-center studies over a longer time period. The additional studies provide more statistical power to detect the effects of potential moderators, such as time and multi- versus single-center trials. Utilizing multivariate procedures allows for simultaneous modeling of both percent days abstinent (the inverse of percent drinking days) and relapse to heavy drinking outcomes, and thus provides independent estimates for each outcome. This procedure is advantageous as it considers the correlation between the modeled outcome variables, instead of assuming independence as two separate univariate procedures do, thereby reducing the likelihood of Type I error (i.e., erroneously rejecting the null hypothesis).

We hypothesize that naltrexone effect sizes continue to decrease after 2003 (the last year examined by Feinn and Kranzler), but eventually plateau (we expect that at no point in time will effect sizes average zero or less) and that multi-center trials continue to yield smaller effect sizes than single-site trials. Further extending the work of Feinn and Kranzler, we planned to examine whether placebo group improvement has been increasing over time, thus accounting for smaller medication effects over time. We also test whether the presence of a placebo run-in period moderates naltrexone effect sizes. Placebo run-in periods eliminate non-compliant participants or participants who exhibit a strong initial placebo response. Thus, studies with a placebo run-in period are expected to have stronger medication effects. Lastly, we examine the interaction between publication year and multi-center (vs. single site), to determine if there are differential effect sizes in multi-center sites versus single site studies on average (grand mean centered intercept) and over time (slope coefficient).

Materials and Methods

Literature Search

Forty-one of the 48 trials examining naltrexone compared with placebo in Maisel et al (in press) reported data to compute ESs for one or both outcomes (percent days abstinent and relapse to heavy drinking) examined in this meta-analysis. Double-blind, placebo-controlled, randomized trials 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 the search term “Naltrexone,” terms for alcohol problems and dependence (e.g., “alcohol*,” “problem drinking”) and terms for randomized controlled trials (e.g., “randomized controlled,” “clinical trial”). In addition to the double-blind, placebo-controlled randomized trial criterion, 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. Further details on the search procedures can be found in Maisel et al. (in press).

Moderator Coding

Several study characteristics (i.e., moderators) were examined to determine if they could explain some or all of the between-study variability in ESs and any effect of time on decreasing effect sizes. Each moderator was double-coded and consensus was reached between the coders and the Project Coordinator (NM).

- *Publication year* ($k=41$ studies) was coded as a continuous variable and indicates the year of study publication.
- *Multi-center design* ($k=41$) was based on whether the study was conducted at one site or at multiple sites (single-site=0 [$k=30$], multi-center=1 [$k=11$]).
- *Placebo run-in* ($k=41$) was coded as a categorical variable and indicates whether a placebo medication was administered to all participants [$k=32$] or not [$k=9$] prior to start of active naltrexone medication.
- *Placebo group improvement* was operationalized as the difference between baseline and end-of-treatment scores on the outcomes, with effect sizes calculated as described below.

Calculation of Effect Sizes

The effect size, Hedges' (2009) g , was computed for (1) percent days abstinent (the inverse of Feinn and Kranzler's 'percentage of days drinking' outcome), but scored in the same direction so that positive effect sizes indicate greater percent days abstinent or a lower percentage of days drinking and (2) relapse to heavy drinking for (a) post-treatment, between-group (naltrexone vs. placebo) means or proportions; and (b) pre-post within-group mean change for the placebo group. The details for computing g can be found in Cooper, Hedges, and Valentine (2009). Note that the study reported by Volpicelli (1995) included additional participants to those focused on in Volpicelli, Alterman, Hayashida, and O'Brien (1992). Although the Volpicelli et al. (1992) study has been cited as the primary study in most previous reviews, including in Feinn and Kranzler, we used data from Volpicelli (1995) to calculate ESs because it presents results for a larger sample.

Statistical Analyses

All analyses involved the simultaneously modeling of effect sizes on the two outcomes (percent days abstinent and relapse to heavy drinking) using a multivariate meta-analytic procedure. A random effects estimator was generated, which assumes that the studies in a meta-analysis were randomly sampled from a population of studies. All meta-analytic steps were conducted using R statistical software packages 'compute.es' (Del Re 2012), 'MAd' (Del Re and Hoyt, 2010) or 'mvmeta' (Gasparrini, 2011).

Simultaneously modeling multiple outcomes using multivariate meta-analytic methods is more statistically appropriate than conducting separate univariate analyses on each outcome.

The univariate approach assumes within-study outcomes are independent (uncorrelated). Violation of this assumption, which is common, generally leads to biased parameters and overestimated ESs (Kirkham et al., 2012; Riley, 2009). Multivariate procedures yield separate estimates for each outcome that account for the covariance between within-study ESs and therefore do not violate the statistical assumption of independence. These procedures have been shown to generate less biased estimates and standard errors, particularly when some included studies do not provide data for both outcomes (Kirkham et al. 2012; Riley 2009). Specifically, the multivariate method maximizes the available data and uses the correlation between outcomes to ‘borrow strength’ (Kirkham et al. 2012) from each non-missing data point to more precisely estimate the population parameters. The formulae and further rationale for multivariate meta-analytic models are detailed elsewhere (Jackson et al., 2011; Kirkham et al., 2012; Riley, 2009).

Heterogeneity in model estimates 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. Q has an approximate χ^2 distribution with $k - 1$ degrees of freedom, where k is the number of studies aggregated. Q -values above the critical value result in rejection of the null hypothesis of homogeneity (Cooper et al. 2009)

Results

Unconditional model

The overall effect in the between-group omnibus analysis (total number of studies $k=41$; total ESs=61) for percent days abstinent was $g_+ = .143$ (95% CI = 0.080, 0.205) and for relapse to heavy drinking was $g_+ = .247$ (95% CI = 0.158, 0.336), indicating a statistically significant between-group difference on each of the two outcomes that favored naltrexone treatment over placebo. Although the overall between-group effect sizes were significant, there was substantial heterogeneity in effect sizes ($Q = 54.79$, $df=38$, $p < .038$; $I^2 = 30.6\%$; see the two Omnibus “Intercept” rows in Table 1 for details that indicate substantial variation in effect sizes across studies) for each outcome when modeled together. We therefore conducted moderator analyses to determine if study characteristics could explain the variability in between-group naltrexone effects.

Single-moderator analyses

Results of the single-predictor moderator analyses are presented in Table 1. The effect of publication year was significant, with naltrexone ESs decreasing over time (see Figure 1). The effect size on percent days abstinent declined by an estimated .022 per year, whereas the effect size for relapse to heavy drinking declined by an estimated .037 per year. The inclusion of a quadratic term to test if declining effect sizes eventually plateaued was not significant, so it was removed from the analyses. Although not statistically significant, placebo group improvement was modestly correlated with smaller naltrexone effect sizes ($r = -0.219$, $p=.28$). Contrary to our hypotheses, multi-center studies did not have smaller ESs than single-site studies and studies with placebo run-in periods did not have larger effect sizes than those without a placebo run-in period.

Multiple-moderator analysis

With the exception of placebo-group improvement, all moderators, as well as the interaction between multi- versus single-center study and year of publication, were entered simultaneously as predictors of effect size in a multivariate, multiple-moderator, meta-regression. Placebo-group improvement was not included in this multiple-moderator analysis because, as described below, only 21 of the 41 studies provided data on this moderator. As shown in Table 2, the multiple-moderator model explained significant

heterogeneity in effect sizes ($Q(30) = 36.98$; $p = .178$; $I^2 = 18.9\%$), while modeling both outcomes simultaneously. However, only year of publication was a significant moderator in the analysis for each outcome. Specifically, effect sizes on percentage of days abstinent were estimated to decrease by $\gamma_1 = -.023$ ($p = .027$) per year, whereas relapse to heavy drinking effect sizes (reversed-scored so that larger effect sizes reflected better outcome) decreased by an estimated $\gamma_2 = -0.054$ ($p < .001$) per year. Neither multi- versus single-site study, the interaction between multi- versus single-site study and year of publication, nor placebo run-in period was a significant moderator in the multiple-moderator models. Thus, although naltrexone ESs relative to placebo have been decreasing over time, this effect cannot be explained by the other moderators examined.

Pre-post improvement in medication and placebo groups

Due to a failure to report pre-treatment values or provide results of significance tests of pre-post change, 17 studies could not be included in the analyses of within-group improvement. Three additional studies were excluded from the analyses because they were outliers that exerted undue influence on the meta-regression models, leaving a total of 21 studies. Within this sample of 21 studies, publication year was not significantly related to naltrexone-placebo between-group effect sizes on percent days abstinent or relapse to heavy drinking. Therefore, we did not examine placebo group improvement as a moderator of effect sizes over time as had been planned.

Discussion

This random-effects, multivariate meta-analysis extends Feinn and Kranzler's (2005) review by examining the efficacy of naltrexone in a larger sample of both single- and multi-site studies and by using multivariate procedures to model independent medication effects on both percent days abstinent and relapse to heavy drinking simultaneously. Each of the findings will be discussed as they relate to the literature on naltrexone pharmacotherapy trials.

Smaller naltrexone effects over time

We found small, but significant effects of naltrexone relative to placebo on both outcomes, although, consistent with its presumed mechanism of action, naltrexone had a larger effect in reducing heavy drinking than in promoting abstinence (see also Maisel et al., in press). However, naltrexone effect sizes have attenuated over time (i.e., more recent studies have shown smaller effects of naltrexone relative to placebo) on both outcomes, but a test of whether decreasing effect sizes have eventually leveled-off (by including a quadratic term for year of publication) was not significant. The linear effect of publication year remained significant even after controlling for several additional moderators, indicating that they did not account for decreases in ESs over time. These results are in partial accord with those of Feinn and Kranzler (2005) who found a significant decrease in naltrexone effects on relapse to heavy drinking over time, but a non-significant decrease in the medication's effects on percent days drinking. This difference may be due to Feinn and Kranzler's study being underpowered ($k=15$ studies) to detect significant effects and/or due to our use of a multivariate versus their univariate procedure. In a post-hoc analysis, the correlations of several moderators speculated to be associated with publication year (severity of patient sample, presence of psychosocial treatment, and percentage of participants not followed) were tested and found to be not significant. Therefore, these additional moderators do not seem to account for the decreasing naltrexone effect sizes over time.

Multi-center versus single-site studies

Contrary to our hypothesis, multi-center studies did not have smaller effects on either drinking outcome relative to single-site studies. These findings are somewhat discrepant with those of Feinn and Kranzler, who found significantly smaller effects for naltrexone in multi-center studies on percent days abstinent, but a non-significant effect on effect sizes for relapse to heavy drinking. The difference between their and our findings again may be due to differences in methods used (a univariate versus a multivariate procedure) or to greater heterogeneity in each group in our analyses, given the additional studies published between 2004 and 2009 that were included.

Interaction between time and single- versus multi-site studies

Feinn and Kranzler noted that multi-center studies may have smaller effects due to variability in study execution among centers (e.g., varying fidelity to treatment protocol, recruitment procedures) and in other contextual factors, such as participant and provider characteristics, promoting variability in medication effects across sites. One speculation is that more recent studies may have had tighter protocols and recruitment procedures which resulted in smaller differences between single- and multi-center studies. To test this, we examined the interaction between publication year and multi-center (vs. single-site) to determine if there were differential effect sizes in multi-center sites versus single site studies over time. However, this analysis yielded a non-significant effect, indicating there were no differences in the effect sizes for single- versus multi-site studies over time.

Placebo run-in

In contrast to our hypothesis, the use of placebo run-in period was not a significant moderator of naltrexone effects. That is, study designs that include a run-in period prior to the start of the trial (to eliminate non-compliant participants or those who exhibit a strong initial response to placebo) did not have larger effect sizes than studies without a run-in period. Placebo run-in periods also did not account for the decrease in naltrexone ESs over time.

Limitations

The primary limitation of this review is related to meta-analytic methods. We examined naltrexone efficacy at the study level and relied on summary statistics from each study, which may obscure more nuanced findings. Further, factors other than the moderators examined here may account for some of the variation in effect sizes, including why naltrexone effect sizes have been decreasing over time.

Conclusion

Overall, the aggregated effect of naltrexone is modest, but significant, relative to placebo in pharmacotherapy trials of alcohol use disorders. However, naltrexone effects have attenuated over time. We were unable to examine whether placebo group improvement explained any of the decline in naltrexone effects over time because, in the subset of studies with placebo group improvement data, naltrexone effects did not decline over time. Contrary to our hypotheses, the moderators that were examined, multi- versus single-site trials and use of a placebo run-in period, did not account either for the effects of naltrexone or for the decrease in naltrexone effects over time. Thus, factors that explain the decreasing effects of naltrexone over time remain to be determined.

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.

References

- Del Re, AC. R package version 0.3. Palo Alto, CA: 2012. Compute.es: Compute Effect Sizes [statistical software program].
- Cooper, HM.; Hedges, LV.; Valentine, JC. The handbook of research synthesis and meta-analysis. Russell: Sage Foundation Publications; 2009.
- Del Re AC, Maisel MC, Blodgett JC, Wilbourne PL, Finney JW. Placebo group improvement in trials of pharmacotherapies for alcohol use disorders: A multivariate meta-analysis examining change over time. *J Clin Psychopharmacol*. (In press).
- Feinn R, Kranzler HR. Does Effect Size in Naltrexone Trials for Alcohol Dependence Differ for Single-Site vs. Multi-Center Studies? *Alcohol Clin Exp Res*. 2005; 29:983–988. [PubMed: 15976524]
- Finney JW. Regression to the mean in substance use disorder treatment research. *Addiction*. 2008; 103:42–52. [PubMed: 17999707]
- Gasparrini, A. R package version 0.2.4. London, UK: 2011. mvmeta: multivariate meta-analysis and meta-regression [statistical software program].
- Harris A, Kivlahan DR, Bove T, Humphreys KN. Pharmacotherapy of alcohol use disorders in the Veterans Health Administration. *Psychiatr Serv*. 2010; 61:392–398. [PubMed: 20360279]
- Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. *Stat Med*. 2011; 30:2481–2498.
- Kemp AS, Schooler NR, Kalali AH, Alphas L, Anand R, Awad G, Davidson M, Dubé S, Ereshefsky L, Gharabawi G, Leon AC, Lepine JP, Potkin SG, Vermeulen A. What is causing the reduced drug-placebo difference in recent schizophrenia clinical trials and what can be done about it? *Schizophr Bull*. 2010; 36:504–509. [PubMed: 18723840]
- Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. *Stat Med*. 2012; 31:2179–2195. [PubMed: 22532016]
- Del Re, AC.; Hoyt, WT. R package version 0.8. Palo Alto, CA: 2010. MAAd: Meta-analysis with mean differences [statistical software program].
- Maisel MC, Blodgett JC, Wilbourne PL, Humphreys K, Finney JW. Meta-analysis of naltrexone and acamprosate for treating alcohol use disorders: When are these medications most helpful? *Addiction*. (In press).
- Rief W, Nestoriuc Y, Weiss S, Welzel E, Barsky AJ, Hofmann SG. Meta-analysis of the placebo response in antidepressant trials. *J Affect Disord*. 2009; 118:1–8. [PubMed: 19246102]
- Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. *J R Stat Soc Series A Stat Society*. 2009; 172:789–811.

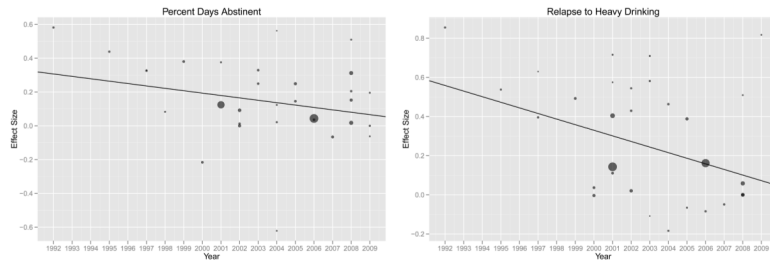


Figure 1.

Note. Hedges' g effect size is on the y -axis and year of publication is on the x -axis for each outcome. Each point represents a study and the size of the point represents the study weight (inverse of variance), where larger points are larger sample size studies and are therefore more precise estimates of the population effect size.

Table 1

Multivariate Omnibus and Single-Moderator Analyses

	<i>k</i>	Intercept	γ_1	95% CI	<i>p</i>
Percentage Days Abstinent					
Omnibus	31	0.143		[0.080, 0.205]	<.001*
Publication Year	31	0.386	-0.022	[-0.040, -0.003]	.021*
Multi-Center Design	31	0.221	-0.110	[-0.276, 0.057]	.197
Placebo Run-In	31	0.159	0.066	[-0.168, 0.299]	.581
Relapse to Heavy Drinking					
Omnibus	30	0.247		[0.158, 0.336]	<.001*
Publication Year	30	0.640	-0.037	[-0.013, -0.060]	.002*
Multi-Center Design	30	0.322	-0.120	[-.354, 0.118]	.328
Placebo Run-In	30	0.266	0.035	[-0.293, 0.363]	.834

Note. Publication year was centered at the first publication in this analysis (1992). Multi-Center Design is coded dichotomously where 1 = multi-site and 0 = single-site. Placebo Run-In was coded dichotomously where 0 = No Run-In and 1 = Yes. Multivariate analyses used a mixed model (studies random, levels of moderator variables fixed);

* *p*-value < 0.05. *k* = number of studies, γ_1 = slope.

Table 2

Multiple-Moderator Multivariate Analyses (k=41; total ESs = 61)

	<i>y</i>	95% <i>CI</i>	<i>p</i>
Percentage Days Abstinent			
Intercept	0.197	[0.070, 0.316]	<.003*
Publication Year	-0.028	[-0.052, -0.003]	.027*
Multi-Center Design	-0.090	[-0.249, 0.069]	.270
Placebo Run-In	-0.061	[-0.299, 0.177]	.615
Pub Year X Multi-Center	0.040	[-0.012, 0.091]	.131
Relapse to Heavy Drinking			
Intercept	0.304	[0.166, 0.441]	<.001*
Publication Year	-0.054	[-0.081, -0.026]	.001*
Multi-Center Design	-0.124	[-0.303, 0.055]	.180
Placebo Run-In	-0.280	[-0.589, 0.033]	.079
Pub Year X Multi-Center	0.050	[-0.011, 0.102]	.120

Note. Overall tests of model significance were $Q(31) = 36.98$, $p = .178$; Publication year was grand mean centered. Multi-Center Design is coded dichotomously where 1 = multi-site and 0 = single-site. Placebo Run-In was coded dichotomously where 0 = No Run-In and 1 = Yes. Multivariate analyses used a mixed model (studies random, levels of moderator variables fixed);

* p -value < 0.05.