



ARTICLE

Do behavioral pharmacology findings predict clinical trial outcomes? A proof-of-concept in medication development for alcohol use disorder

Lara A. Ray^{1,2}, Han Du¹, ReJoyce Green¹, Daniel J. O. Roche¹ and Spencer Bujarski¹

Behavioral pharmacology paradigms have been used for early efficacy testing of novel compounds for alcohol use disorder (AUD). However, the degree to which early efficacy in the human laboratory predicts clinical efficacy remains unclear. To address this gap in the literature we employed a novel meta-analytic approach. We searched the literature for medications tested for AUD using both behavioral pharmacology (i.e., alcohol administration) and randomized clinical trials (RCTs). For behavioral pharmacology, we computed medication effects on alcohol-induced stimulation, sedation, and craving during the alcohol administration ($k = 51$ studies, 24 medications). For RCTs, we computed medication effects on any drinking and heavy drinking ($k = 118$ studies, 17 medications). We used medication as the unit of analysis and applied the Williamson-York bivariate weighted least squares estimation to preserve the errors in both the independent and dependent variables. Results, with correction for publication bias, revealed a significant and positive relationship between medication effects on alcohol-induced stimulation ($\beta = 1.18$, $p < 0.05$), sedation ($\beta = 2.38$, $p < 0.05$), and craving ($\beta = 3.28$, $p < 0.001$) in the laboratory, and drinking outcomes in RCTs, such that medications that reduced stimulation, sedation, and craving during the alcohol administration were associated with better clinical outcomes. A leave-one-out Monte Carlo analysis examined the predictive utility of these laboratory endpoints for each medication. The observed clinical effect size was within one standard deviation of the mean predicted effect size for all but three pharmacotherapies. This proof-of-concept study demonstrates that behavioral pharmacology endpoints of alcohol-induced stimulation, sedation, and craving track medication effects from the human laboratory to clinical trial outcomes. These results apply to alcohol administration phenotypes and may be especially useful to medications for which the mechanisms of action involve alterations in subjective responses to alcohol (e.g., antagonist medication). These methods and results can be applied to a host of clinical questions and can streamline the process of screening novel compounds for AUD. For instance, this approach can be used to quantify the predictive utility of cue-reactivity screening models and even preclinical models of medication development.

Neuropsychopharmacology (2020) 0:1–9; <https://doi.org/10.1038/s41386-020-00913-3>

INTRODUCTION

Behavioral pharmacology has a long and rich history in addiction science, from examining drug effects under controlled laboratory conditions to testing risk mechanisms for alcohol and other substance use disorders [1, 2]. More recently, behavioral pharmacology approaches have been proposed as tools for medication development for addiction [3–6], with the most commonly used paradigm consisting of controlled alcohol/drug administration (i.e., alcohol or drug “challenge”). Controlled drug/alcohol challenges allow for tests of medication \times drug/alcohol interactions, which is critical in medication development for establishing safety and tolerability. Beyond safety, behavioral pharmacology paradigms permit testing of theoretically meaningful endpoints, described as early efficacy markers. These endpoints often include medication-induced changes in the subjective response (SR) to alcohol or drugs, as well as measures of craving, via cue-reactivity or alcohol/drug administration [7–9]. Putatively, these early efficacy endpoints (i.e., Phase Ib) can inform clinical trials and whether or not a novel compound should be

advanced to the next stage of clinical testing (i.e., Phase II, randomized clinical trial) [6, 10]. However, the human laboratory prediction made herein is limited to behavioral pharmacology studies with alcohol administration. Furthermore, certain medications (e.g., antagonists such as naltrexone) may be better suited for screening through these models than other medications (e.g., antagonist such as gabapentin).

While the utility of behavioral pharmacology for establishing the safety and tolerability of addiction pharmacotherapies in humans is well established, the degree to which the early efficacy of novel compounds in the human laboratory can predict clinical efficacy remains unclear. In a recent critique, we argued that the degree to which a behavioral pharmacology paradigm is useful as an early efficacy marker depends on the degree to which that paradigm is related to the desired clinical outcome (e.g., abstinence or reduced heavy drinking) [11]. At a theoretical level, medications that reduce alcohol-induced stimulation and alcohol craving during alcohol administration are thought to reduce alcohol use [12]. Medications that potentiate the sedative effects

¹University of California, Department of Psychology, Los Angeles, CA, USA and ²University of California, Department of Psychiatry and Biobehavioral Sciences, Los Angeles, CA, USA

Correspondence: Lara A. Ray (lararay@psych.ucla.edu)

Received: 25 July 2020 Revised: 31 October 2020 Accepted: 5 November 2020

Published online: 24 November 2020

of alcohol during alcohol administration are also thought to reduce alcohol intake [12]. Nevertheless, these hypothetical predictions have not been empirically tested and doing so is the focus of the present study. In our previous work, we conducted a series of simulations were conducted to determine the required sample size for a behavioral pharmacology study to detect early efficacy based on varying levels of association between human laboratory paradigms and clinical outcomes [11]. These simulations used hypothetical associations, given that estimates of the “real” association between medication effects in behavioral pharmacology studies and its efficacy in randomized clinical trial (RCTs) remain unknown.

Thus far, systematic reviews have focused on qualitative assessments of the consistency between behavioral pharmacology and RCT results [3, 13, 14]. For example, Naltrexone has known clinical efficacy for AUD [15, 16] and appears to reliably blunt the reinforcing effects of alcohol [17, 18]. This is seen as evidence that reducing the rewarding effects of alcohol is a mechanism of action of naltrexone. Furthermore, blunting the rewarding SR to alcohol is an early efficacy marker of naltrexone observed in the lab [19, 20] and confirmed in clinical trials [21, 22]. Though these reviews provide insights into the consilience between behavioral pharmacology and clinical outcomes, they cannot provide quantitative estimates that could inform medication development. To date, no quantitative test of the concordance of behavioral pharmacology and clinical efficacy has been published in the field of addiction or psychiatry.

To address this gap in the literature we employed a novel translational meta-analytic approach to test whether behavioral pharmacology effect sizes are correlated with RCT effect sizes. To accomplish this goal, we searched the literature for medications tested for AUD using both behavioral pharmacology paradigms and RCTs. We then computed effect sizes for each medication in both behavioral pharmacology (i.e., “lab”) and clinical trial (i.e., “clinic”) designs. To integrate these two independent effect sizes, we used medication as the unit of analysis and tested the degree to which behavioral pharmacology effects are correlated with treatment effects across the various AUD pharmacotherapies.

For this proof-of-concept study, we focus on the primary outcome from alcohol challenge studies, which is SR to alcohol, including alcohol craving. This study does not address other important early efficacy endpoints in behavioral pharmacology for AUD, including cue-induced craving [23] and stress-induced [9, 24] craving. The focus on SR is consistent with its centrality in multiple theories of AUD etiology [25–28], and its relevance and wide prevalence in the AUD behavioral pharmacology literature [4, 12, 26]. Furthermore, alcohol craving has been introduced as an AUD symptom in DSM-5 and is widely considered a translational phenotype [29]. Through independent meta-analysis of AUD pharmacotherapies in human laboratory studies and RCTs, and by systematically testing their association, this study will quantitatively estimate the relationship between behavioral pharmacology and clinical trial outcomes in medication development for AUD.

METHODS

Literature review

Inclusion criteria for the behavioral pharmacology studies were (1) the administration of a pharmacological agent approved or being developed for the treatment of AUD, (2) alcohol administered in the laboratory to a target BrAC via alcohol challenge or priming for self-administration¹, (3) SR outcomes measured via self-report

questionnaires, (4) reported in the English language, or translated to English, and (5) publication in a PubMed indexed journal. Databases were searched through July, 2018 and collected data were analyzed through September, 2020.

Given the scope of literature covered in this meta-analysis an algorithmic approach was utilized to identify all the relevant research reports. First, published reviews of AUD psychopharmacology were reviewed to identify medications that have been tested in the human laboratory with alcohol administration paradigms [3, 4, 30–32]. Examination of published reviews identified 40 pharmacological compounds that may have been evaluated using behavioral pharmacology paradigms from 45 laboratory studies. Second, PubMed searches were conducted with each of the 40 medications in combination with any of the following phrases: “alcohol challenge,” “alcohol response,” “response* to alcohol,” “alcohol response,” “alcohol priming,” “alcohol intoxication,” “ethanol intoxication,” “response* to ethanol,” “ethanol response.” Medical subject headings (MeSH) were used in combination with terms listed above. These PubMed searches yielded a total of 1206 studies which were assessed for relevance in the present paper via abstract review.

From these 1206 initial studies, 67 were deemed relevant for full text review. 16 studies were excluded based on full text review (7 for lack of controlled alcohol administration, 4 for lack of SR outcomes, 2 for lack of new outcomes, and 3 for self-administration only). This resulted in a final sample of 51 studies that were included in this analysis comprising 55 independent samples with 1850 total subjects (all study statistics are made publicly available in <https://github.com/sbujarski>). All studies were coded by at least two raters (SB, RG, and/or DJOR). Where coding discrepancies existed, all raters met in person to reach a consensus. Furthermore, when sufficient data to generate effect size estimates were not reported in the published paper, corresponding authors were contacted via email to obtain the necessary information. The Digitizelt software [33] was also utilized to extract data from published figures [34].

Inclusion criteria for the RCT studies was: (1) a randomized controlled trial, (2) double or single blinded, (3) Placebo or active control condition, (4) Alcohol use was the primary endpoint, (5) 4 or more weeks of treatment, and (6) 12 or more weeks of follow-up. These inclusion criteria were selected based on established guidelines by the Cochrane Collaboration. Similar to the behavioral pharmacology review, RCT literature searching was algorithmic. First, Cochrane reviews were searched on each of the 24 medications with behavioral pharmacology data. Six medications (Naltrexone, Nalmefene, Acamprosate, Topiramate, Gabapentin, and Zonisamide) had published Cochrane reviews for AUD which included a total of 67 studies. Secondly, PubMed searches were conducted on each of the 24 medications with the following search phrases: “randomized clinical trial,” “randomized controlled trial,” “randomised clinical trial,” “treatment,” and “Alcohol.” For medications that had Cochrane reviews, PubMed searches were time frame restricted to two years prior to the publication of the Cochrane review to the present. These searches identified a total of 2028 records, 132 of which were new studies subjected to full-text review and 118 which were included in the analyses. For RCTs, there were 17 medications and the number of studies for each medication varied from 1 to 34. The systematic review process is shown in Fig. 1.

Selection of outcomes

Behavioral pharmacology. Prior factor analytic work by our group suggested that SR to alcohol represents a multifaceted construct with four distinct domains: (a) Stimulation/Hedonia, (b) Craving/Motivation, (c) Sedation/Motor Intoxication, and (d) Negative Affect [35, 36]. Assigning of outcome variables to SR domains was determined through consensus discussion among all study coders referencing the prior factor analytic work [35, 36], other published articles, and/or through referencing the specific items. The specific

¹Studies that only reported subjective response data in the context of a self-administration paradigm were excluded due to the potential for large confounding effects of BrAC differences between medication groups.

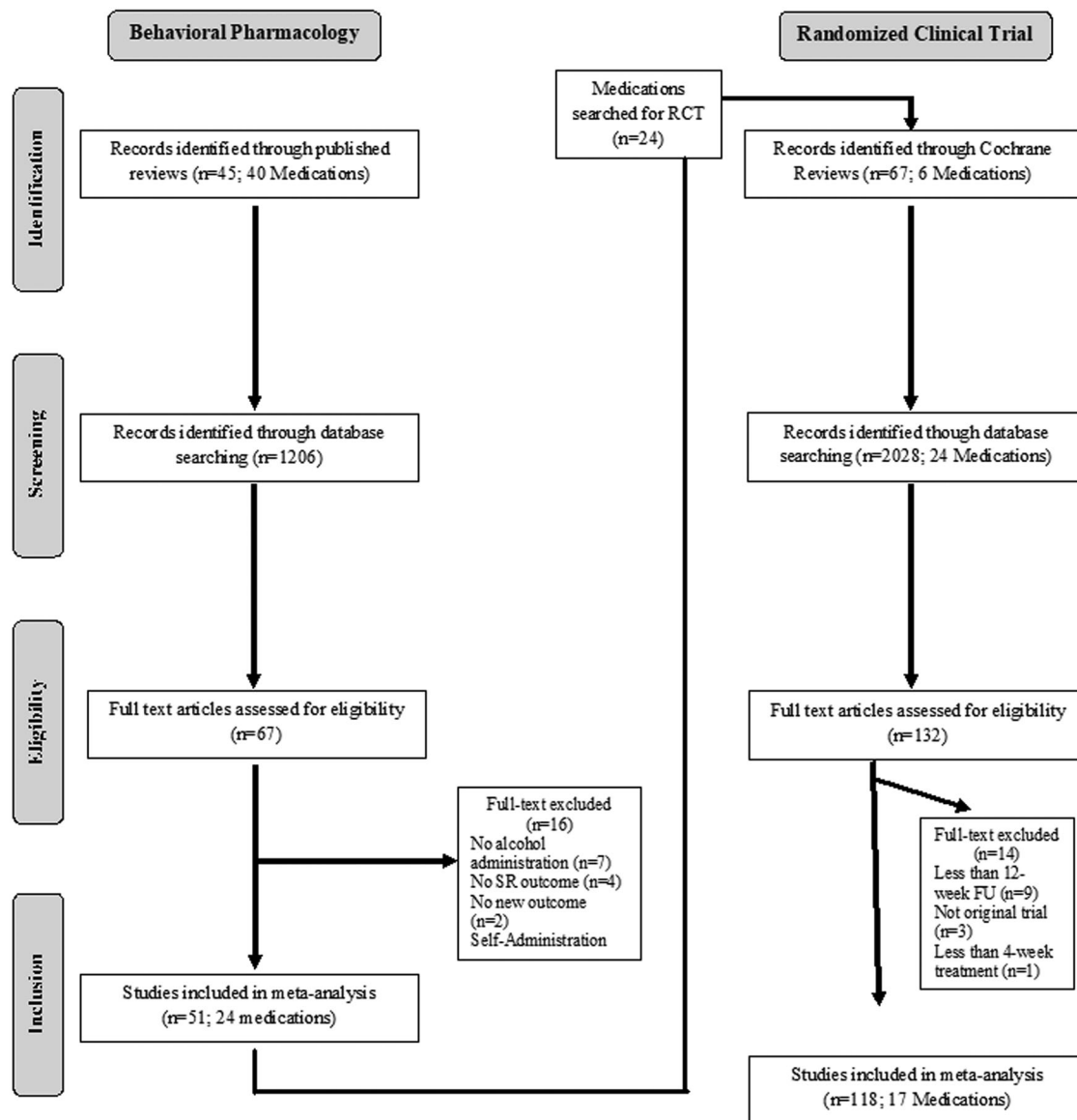


Fig. 1 Consists of a flow chart of the systematic review process. Unlike a traditional meta-analysis, the goal was to identify trials for AUD medications tested in behavioral pharmacology and to restrict the analyses of randomized clinical trials to only those medications studied under behavioral pharmacology involving alcohol administration.

domain assignments are presented in Supplementary Fig. 1. Separate meta-analyses of medication effects were conducted on each outcome domain.

Randomized clinical trials. Informed by FDA guidelines for AUD medication development [37], two types of RCT outcomes were analyzed: any drinking and heavy drinking. For heavy drinking, the continuous outcome of percent drinking days (or percent heavy drinking days) was analyzed. The two clinical outcomes for RCTs were combined into a single outcome. This approach is consistent with standard practice in meta-analysis for AUD/SUD [38, 39], and resulted in a more stable estimate of medication effects on alcohol use while reducing the number of independent tests/comparisons. Meta-analytic methods for the RCT studies were identical to those employed for the behavioral pharmacology.

Data analytic plan

Data analysis for this study consisted of several steps. First, we calculated the unbiased Cohen's *d* as the target effect size for each

study. Cohen's *d* was defined as the mean from the treatment group minus the mean from the control group divided by a pooled standard deviation. Cohen's *d* was corrected by multiplying a correction factor to obtain an unbiased Cohen's *d*. Second, we grouped the effect size results from Abstinent and Heavy Drinking together. The effect sizes of Abstinent and Heavy Drinking were in the opposite directions, therefore we reverse-coded the effect sizes of Abstinent. After reverse-coding, in both Abstinent and Heavy Drinking, a negative effect size indicates that the treatment group has a lower group mean than the control group. Hence, there are 4 outcomes in behavioral pharmacology laboratory (Stimulation/Hedonia, Craving/Motivation, Sedation/Motor Intoxication, and Negative Affect) and 1 outcome (Abstinent and Heavy Drinking were combined) in clinical trials. Third, within each outcome, we conducted fixed-effects meta-analysis for each medication using the *metaphor* R package [40]. In other words, all medications studies identified in our literature search were coded for their effects on the four behavioral pharmacology outcome domains and the single clinical trial

outcome domain. And the effects of each study were pooled into a single estimate for a given medication. Fixed-effects meta-analysis was used instead of random-effects meta-analysis because for some medications, there was only 1 or 2 studies. In this case, we do not have enough studies to accurately estimate both the overall effect size and between-study heterogeneity. Hence, we adopted the fixed-effects meta-analysis and estimated the overall effect size only. For Stimulation, there were 17 medications. Within each medication, the number of studies varied from 1 to 17. For Sedation, there were 20 medications. Within each medication, the number of studies varied from 1 to 18. For Craving, there were 17 medications. Within each medication, the number of studies varied from 1 to 16. For Negative Affect, there were only 8 medications. Within each medication, the number of studies varied from 1 to 7. Since there were only a few studies for Negative Affect and data information was sparse, we excluded Negative Affect in the next step. Fourth, we aimed to use the effect size of each medication in the behavioral pharmacology laboratory to predict the effect size of each medication in clinical trials. Considering that both the independent and dependent variables have errors, we used the Williamson-York bivariate weighted least squares estimation to preserve the errors in both the independent and dependent variables [41–44]. The widely used ordinary least squares estimation could not be applied here because it only considers the errors in the dependent variable, and thus important information of the independent variable would be omitted. There were three regressions based on different laboratory outcomes (excluding negative mood due to its low data availability): Stimulation effect sizes predict clinical effect sizes, Sedation effect sizes predict clinical effect sizes, and Craving effect sizes predict clinical effect sizes. Fifth, we conducted a sensitivity analysis by correcting for publication bias. We used the *p-uniform* method [45], obtained the corrected estimated overall effect sizes, and conducted regression analysis. Compared to other publication bias correction methods, the *p-uniform* method performs relatively well when the effect sizes are homogeneous and the sample size is small [46]. We used the *puniform* R package [47].

Subsequent to analyzing the bivariate associations between laboratory and clinical outcomes we conducted predictive analysis to determine the degree to which these methods can inform go/no-go decisions for clinical trials of novel medications. To assess the predictive utility of these laboratory outcomes, we employed novel a leave-one-out Monte Carlo simulation method. The Williamson-York regression models were trained on a dataset with a single medication removed (the target medication). The regression models were then used to predict the clinical effect size of the target medication based on its observed laboratory effect size. A Monte Carlo method was used to account for predictor value uncertainties. Specifically, 100,000 predicted values were generated for each laboratory outcome. These simulated predicted values were then summarized with respect to their mean and standard deviation. To arrive at a single predicted clinical effect size distribution for the target medication, we computed an aggregated mean and SD across different outcomes. To provide a metric for how accurate these predicted effect sizes were, we compute a z-score for the observed clinical effect size with respect to the predicted mean effect size and standard deviation. This metric therefore represents the degree to which the observed effect size is expected under the predicted range. This procedure was then repeated across all medications included in this study.

Together, this novel application of Williamson-York bivariate weighted least squares estimation, derived from physics and astronomy fields, allowed us to integrate decades of research into a meaningful and quantitatively sound test of relationship between independent effect sizes obtained in behavioral pharmacology and RCT contexts. In this effort, medication was the unit of analysis.

The novel leave-one-out Monte Carlo analysis also provides new insights into the predictive utility of these laboratory methodologies that can inform go/no-go decisions for novel medication clinical trials.

RESULTS

Effect size estimation

Effect size estimation across the 51 human laboratory studies included in the study and across the three outcomes of stimulation, sedation, and craving, are presented in Supplementary Fig. 2. All studies are listed by author/year, medication name, medication dosage, estimated effect size of Hedge's G (converted to Cohen's d for the analyses), average drinks per month in the sample (DpM), and Breath Alcohol Concentration (BrAC) during the alcohol challenge. Effect size estimation across the 118 RCTs included and across the two outcomes of abstinence and heavy drinking are presented in Supplementary Fig. 3. All studies are listed by author/year, medication name, medication dosage, estimated effect size of Hedge's G (converted to Cohen's d for the analyses), and treatment duration (in weeks).

Alcohol-induced stimulation and clinical outcomes

As described above, we tested a model in which the stimulation effect sizes predict clinical effect sizes, across all medications studied under both behavioral pharmacology and RCTs. Effect sizes for stimulation and clinical outcomes were available for 12 medications. The slope of the regression was positive and estimated at $\beta = 1.64$ ($SE = 0.46$, $p < 0.01$) when the laboratory outcome was Stimulation, which indicated a significant positive relationship between the effect sizes for medication effects on alcohol-induced stimulation in the behavioral pharmacology studies and the medication effect sizes in clinical trials for AUD; see Fig. 2. The positive relationship suggests that medications that decreased alcohol-induced stimulation in the human laboratory were found to decrease drinking in RCTs. The bivariate-weighted correlation between the two sets of effect sizes is $r = 0.370$. With publication bias correction and corrected effect sizes, the slope of the regression was estimated at $\beta = 1.18$ ($p < 0.05$), such that the conclusion remained the same.

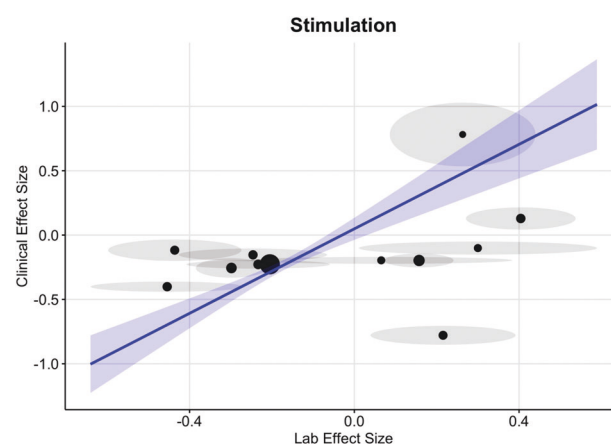


Fig. 2 Displays the Williamson-York bivariate weighted regression in which stimulation effect sizes predict clinical effect sizes. Each medication is represented as a dot on the regression line and smaller dots indicate more error variance while larger dots indicate less error variance around each estimate. Bivariate effect size standard errors for each medication are represented with the ellipses surrounding each point. The regression standard error is represented by the ribbon around the regression line.

Alcohol-induced sedation and clinical outcomes

For the Sedation effect sizes, a positive effect size indicates that the treatment group has a larger effect than the control group, while for the clinical outcomes, a negative effect size indicates that the treatment group has a larger effect than the control group. Data for 13 medications was available.

Results for the model in which the sedation effect sizes predict clinical effect sizes, the slope of the regression was $\beta = 4.04$ ($SE = 2.48$, $p = 0.130$), which was nonsignificant. Correlation between the two sets of effect sizes is $r = 0.227$. With publication bias correction and corrected effect sizes, the slope of the regression was significant and positive, at $\beta = 2.38$ ($p < 0.05$). The significant positive slope indicated that medications which lead to larger increases in sedative subjective effects had poorer clinical benefit see Fig. 3.

Alcohol-induced craving and clinical outcomes

The final model tested whether craving effect sizes predict clinical effect sizes, across all medications studied under both behavioral pharmacology and RCTs. Data was available for 13 medications. The observed slope of the regression was positive and significant, at $\beta = 1.14$ ($SE = 0.32$, $p < 0.01$). This finding suggests that medications that decreased alcohol-induced craving during an alcohol challenge were found to decrease drinking in RCTs. The correlation between the two sets of effect sizes is $r = 0.074$. With publication bias correction and corrected effect sizes, the slope of the regression was $\beta = 3.28$ ($p < 0.001$), such that the significant conclusion remained the same Fig. 4.

Predictive utility of laboratory effects

The leave-one-out Monte Carlo analysis suggested that these the combination of these laboratory and quantitative methods can provide useful information value for predicting clinical efficacy for a novel medication that has yet to be tested in a clinical trial. That said, the effect size uncertainties are generally wide, driven largely by the laboratory effect size precision and modest correlations between laboratory and clinical effects (see Table 1, Fig. 5). The predicted effect sizes were well calibrated and not systematically biased. The average z-score of the observed clinical effect size with respect to the predicted distributions was very small (-0.004). Despite generally high concordance between predicted and observed effects, there were a few medications where

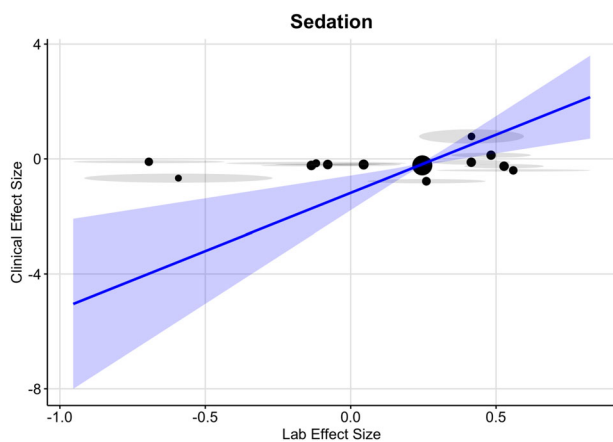


Fig. 3 Displays the Williamson-York bivariate weighted regression in which sedation effect sizes predict clinical effect sizes. Each medication is represented as a dot on the regression line and smaller dots indicate more error variance while larger dots indicate less error variance around each estimate. Effect size standard errors are represented with the ellipses surrounding each point. The regression standard error is represented by the ribbon around the regression line.

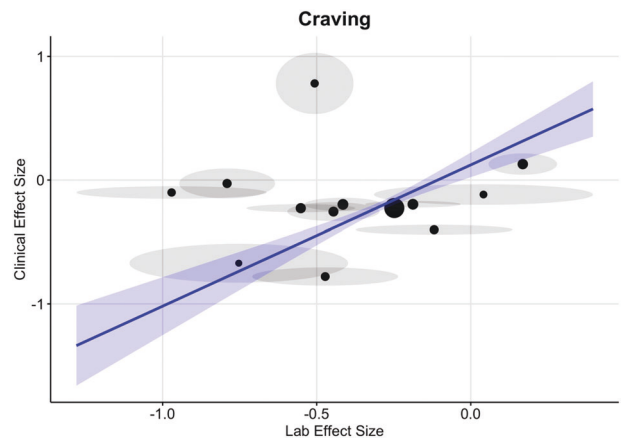


Fig. 4 Displays the Williamson-York bivariate weighted regression in which craving effect sizes predict clinical effect sizes. Each medication is represented as a dot on the regression line and smaller dots indicate more error variance while larger dots indicate less error variance around each estimate. Effect size standard errors are represented with the ellipses surrounding each point. The regression standard error is represented by the ribbon around the regression line.

substantial discrepancies occurred. Namely, Gabapentin was shown to have a significantly larger clinical impact than predicted and Memantine was found to have a significantly more deleterious clinical effect than predicted. Olanzapine was also found to have a smaller clinical impact than predicted, though this effect was substantially less severe than Gabapentin and Memantine. For all other medications the observed clinical effect size was within one standard deviation of the mean predicted effect size.

DISCUSSION

This study tested the relationship between early efficacy assays of SR to alcohol collected in placebo-controlled behavioral pharmacology studies of medications for AUD and the clinical effects of these AUD medications in RCTs. Leveraging advanced meta-analytic tools and the Williamson-York bivariate weighted least squares estimation, the latter appropriate for integrating dependent and independent variables with errors, this proof-of-concept study provided quantitative estimates to a critical substantive question in medication development. Namely, does early efficacy in the human laboratory captured by medication effects on SR to alcohol administration (i.e., stimulation, sedation, and craving) predict clinical outcomes in RCTs for those medications?

Simply put, we predicted that the more a medication reduced alcohol-induced stimulation, relative to placebo, the more that medication reduced alcohol intake in RCTs. This hypothesis was supported by our analyses such that reduced stimulation in the laboratory was positively associated with less drinking in RCTs, across the available medications studied in both human laboratory and clinical settings. Furthermore, we found the same pattern to be true for alcohol-induced craving and sedation, such that reduced craving and sedation in the laboratory was positively associated with less drinking in RCTs. These extensive and innovative analyses across a wide range of medications and outcomes, effectively integrates two critical phases of medication development, namely phase Ib (early efficacy) and phase II (clinical efficacy). It provides critical insights into the degree to which these early efficacy markers (i.e., SR during alcohol challenge) measured in the human laboratory, predict real-world clinical outcomes for AUD in RCTs.

While the fact that there is some consilience across the effects obtained in behavioral pharmacology trials and in RCTs for AUD is

Table 1. Represents the predicted clinical effect size based on each medications' laboratory effect sizes using a leave-one-out Monte Carlo simulation method on the bivariate-weighted regression models.

	Number of laboratory outcomes	Predicted clinical effect size (SD)	Observed clinical effect size	Observed effect size Z-Score
Acamprosate	3	−0.081 (0.293)	−0.196	0.395
Aripiprazole	3	−0.242 (0.373)	−0.117	−0.333
Baclofen	3	−0.433 (1.124)	0.129	−0.499
Gabapentin	3	−0.044 (0.112)	−0.779	6.553
Memantine	3	−0.106 (0.233)	0.782	−3.811
Nalmefene	3	−0.468 (0.337)	−0.228	−0.712
Naltrexone	3	−0.245 (0.261)	−0.227	−0.069
Olanzapine	1	−0.898 (0.706)	−0.028	−1.233
Ondansetron	3	−0.416 (0.316)	−0.256	−0.508
Quetiapine	3	−0.305 (0.482)	−0.101	−0.423
Ritanserin	2	−0.407 (0.449)	−0.154	−0.564
Topiramate	3	−0.166 (0.312)	−0.401	0.755
Varenicline	3	0.046 (0.275)	−0.197	0.885
Zonisamide	2	−4.479 (7.709)	−0.671	−0.494

Predictive model calibration was determined through examining where the observed effect size falls in the predicted effect size distribution through a z-score transform.

encouraging, the magnitude of these associations (i.e., their correlation) was relatively small. As detailed in our simulation study [11], the magnitude of the association between laboratory and clinical outcomes should inform power analyses for human laboratory trials. In that Monte Carlo Simulation study, a correlation between laboratory and clinical outcomes of 0.3 was the smallest and indicated that laboratory studies should have twice the sample size of a clinical trial in order to detect a medium effect size treatment. To further inform go/no-go decisions for novel medications, we conducted a leave-one-out Monte Carlo analysis on the combined human laboratory data. Findings suggested that these the combination of these laboratory endpoints can provide useful information for predicting clinical efficacy for a novel medication that has yet to be subjected to a clinical trial. A caveat to this conclusion is that the effect size uncertainties are generally wide, driven largely by the laboratory effect size precision and modest correlations between laboratory and clinical effects. Furthermore, while there was generally high concordance between predicted and observed effects, there were a few notable exceptions. Specifically, Gabapentin was shown to have a significantly larger clinical impact than predicted and Memantine was found to have a significantly more deleterious clinical effect than predicted. In brief, the Monte Carlo analyses add medication-specific results and directly examine the predictive utility of human laboratory models focused on SR domains.

This study represents an important step toward optimizing the medication development pipeline by leveraging behavioral pharmacology designs to elucidate medication effects on early efficacy endpoints. Insofar as SR to alcohol during an alcohol challenge is the used, and early efficacy endpoints include stimulation, sedation, and craving, this study confirms that these early efficacy markers are indeed quantitatively related to clinical outcomes in RCTs across a range of medications studied under both experimental conditions. In other words, medications that can reduce stimulation, reduce craving, and potentiate sedation during alcohol administration, compared to placebo, fare better in clinical trials as demonstrated by reduced alcohol consumption. This finding is consistent with the role of behavioral pharmacology in early signal detection and screening of promising compounds, as articulated in the medication development literature for AUD [4, 6, 48]. Nonetheless, caution should be exercised in adequately powering studies to reliably detect the behavioral pharmacology

endpoints reported herein. In addition, it is important to consider medication development for AUD and its success, in the broader context of factors, including the lack of substantial investment compared to other fields [49].

During the peer-review process of this study, a number of important caveats were raised and should be considered by the readers in interpreting these findings. These analyses do not distinguish between drugs with a mechanism of action aimed at antagonizing the rewarding effects of alcohol (e.g., naltrexone, nalmefene, topiramate) and medications that seek to maintain abstinence by restoring homeostasis in brain systems dysregulated by the onset of abstinence (e.g., acamprosate and gabapentin). We are clearly underpowered to do so. Nevertheless, it is plausible that the behavioral pharmacology paradigms associated with alcohol administration in the laboratory, and studied herein, may be best suited for testing antagonist medications and less suited for screening the therapeutic potential of medications in the agonist category. Another issue brought up in peer-review is the notion that reduction in heavy drinking may be the ideal primary outcome for an antagonist medication, such as naltrexone [50], whereas abstinence may be a better outcome for an agonist medication, such as acamprosate [51]. In this meta-analysis, abstinence and heavy drinking outcomes are combined in order to boost statistical power. It is plausible that in addition to refining the behavioral pharmacology testing by selecting laboratory outcomes that are best suited based on the mechanism of action of a given medication (i.e., agonist versus antagonist), such refinement should be considered at the level of the clinical outcomes selected.

Several caveats and limitations should be applied to the interpretation of these findings. First, this proof-of-concept study is restricted to three dimensions of SR measured during an alcohol administration paradigm (i.e., stimulation, sedation, and craving). This study does not speak to other important early efficacy endpoints in behavioral pharmacology for AUD, including cue-induced craving [23] and stress-induced [9, 24] craving. Second, this study only examined medications that were studied under both human laboratory and RCT condition when certainly a host of medications did not meet this criterion. Nevertheless, the novel implementation of the Williamson-York bivariate weighted least squares estimation allowed us to integrate independent samples (i.e., participants tested in the laboratory were not the same as

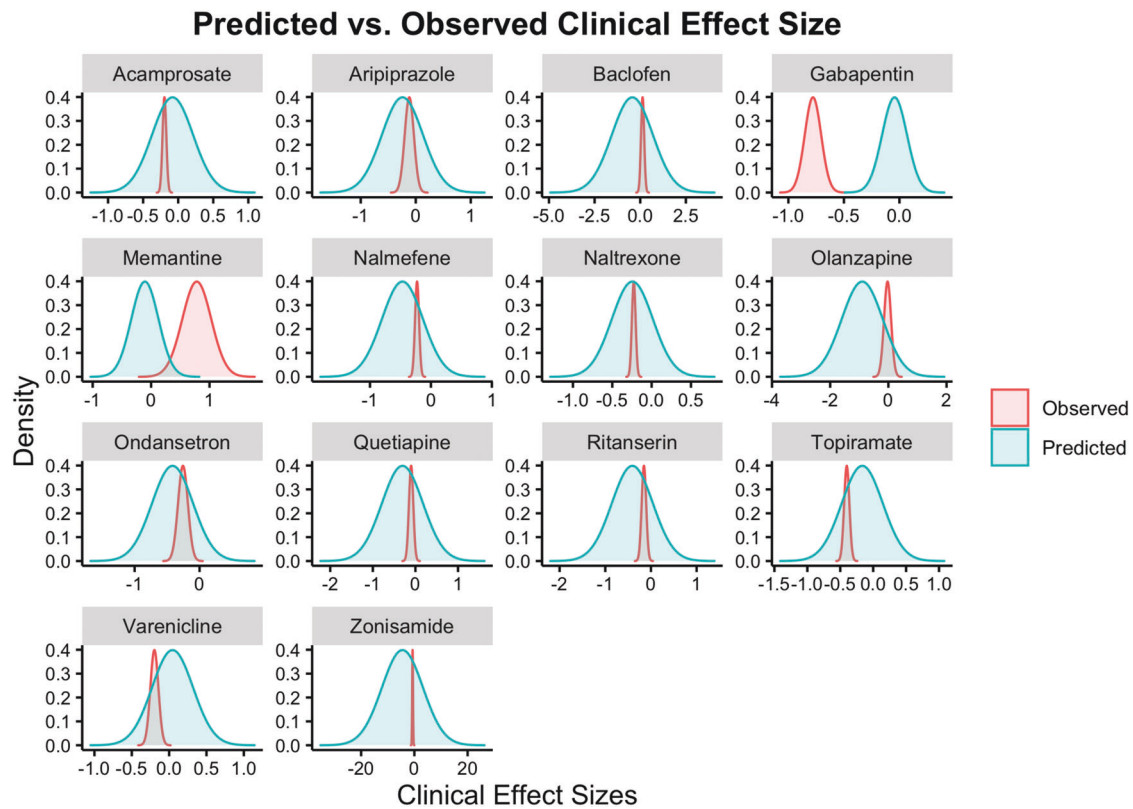


Fig. 5 Displays the predicted and observed clinical effect size distributions. Predicted effect size distributions were generated using a leave-one-out Monte Carlo simulation method for each medication across laboratory effect sizes. Where multiple laboratory effects existed for a given medication, these predicted distributions were aggregated.

those tested in clinical studies). By doing so, we integrated decades of research. The alternative approach would be to test the same participants in the lab before they proceed to a clinical trial [23], which is both costly and cumbersome. Third, utilizing these three early efficacy endpoints to screen novel medications assumes that all promising AUD medications will work through these mechanisms of attenuating craving, stimulation, and/or potentiating sedation during alcohol administration. Conversely, as we understand novel drugs and novel mechanisms of action, a wider range of early efficacy endpoints may be necessary, including assessments of mood, alcohol metabolism, cue-reactivity, and alcohol self-administration [30, 52]. It is plausible that Gabapentin, for example, operates through different mechanisms hence the prediction via SR measures was not consistent with clinical trial outcomes, which proved more favorable clinically than predicted by the model. This is consistent with the argument that agonist medications seeking to restore homeostasis in brain systems dysregulated during abstinence may be better screened through alternative behavioral pharmacology models, including alcohol cue-reactivity, for example. Furthermore, the biobehavioral assays studied herein can inform the development of treatment responsive biomarkers, which remains a critical gap in AUD treatment development [53]. Fourth, publication bias continues to be a problem, and in this study alone, we estimated that 35% of the outcomes mentioned in publications did not have accompanying results. Selective publication of outcomes is endemic in human laboratory studies and clinical research more broadly [54]. While this issue has been recognized for almost three decades [55], it continues to be a threat to the interpretation of scientific findings and to meta-analytic efforts such as ours. Fifth, there is a clear imbalance with regards to the number of studies available across the range of medications studied, clearly naltrexone and acamprosate are the most widely studied medications with

multiple studies available allowing for a more precise estimation of both human laboratory and RCT outcomes. For the other study medications, only a few studies were available for analyses. This imbalance led to more variability in the estimates for studies with few trials and caused medications like naltrexone and acamprosate to exert an undue influence on the outcomes. Nevertheless, since the analyses were conducted with medication as the unit of analysis, then medications with multiple studies were summarized into a single data point such that they did not “count more heavily” in the final analyses than any other medication. Sixth, while these extensive efforts include coding of study covariates, we were not able to reliably implement meta-regression analyses controlling for study differences given that many medications only had a few studies. Additional analyses including covariates may be possible for medications with multiple trials [17]. Seventh, the categorization scheme using items for the dimensions of SR on the basis of their face-validity can be improved upon in future studies in which person-level data are available. Specifically, network analysis may be well-suited for testing the relationships among the predictor variables (i.e., specific items/scales capturing dimensions of SR to alcohol) and in turn, improve the overall model prediction. Eighth, visual inspection of Fig. 5, in which predicted and observed effects are displayed for each medication, suggest that specificity and negative predictive value are low. This means that the lab models studied herein did not correctly identify any medications that were clinically ineffective. However, it should be noted that this is sample of AUD pharmacotherapies that was intentionally selected to have both human behavioral pharmacology and RCT studies. As such, many medications tested in the human lab may have not moved to RCT testing on the bases of poor human-lab outcomes. Ninth, this proof-of-concept study is focused exclusively on medications for AUD and the translation from early efficacy testing (behavioral pharmacology,

phase Ib trial) to clinical efficacy testing (RCT, phase II trial). Nevertheless, the novel methods employed in this study are flexible and can be applied to examining the consilience between preclinical efficacy and early efficacy or clinical efficacy, another longstanding gap in the literature [56, 57]. This approach could also be used to estimate the utility of a host of paradigms for screening medications for alcohol and drug use disorders (e.g., cue-induced craving and self-administration) [5, 58].

In sum, behavioral pharmacology endpoints of alcohol-induced stimulation, sedation, and craving track medications effects from the human laboratory to clinical trial outcomes. This proof-of-concept study uses a novel methodological approach to integrate decades of medication development research and to demonstrate the relationship, albeit of small-to-moderate magnitude, between behavioral pharmacology with alcohol administration and clinical trials endpoints for AUD. These methods and results can be applied to a host of clinical questions and can streamline the process of screening novel compounds for AUD. This methodological approach can be used to quantify the predictive utility of cue-reactivity screening models and even preclinical models of medication screening.

FUNDING AND DISCLOSURE

Support for data analysis and manuscript preparation provided by K24AA025704. The funder had no role in the design, analysis, interpretation, or writing of the report. None of the authors have any competing financial interest in relation to this work. None of the authors have any conflict of interests.

AUTHOR CONTRIBUTIONS

SB and LAR conceptualized and designed the study. SB, DJOR, and RG conducted literature searches and coded all studies. SB and HD conducted all the study analyses. All authors contributed to the interpretation of the data. LR drafted the manuscript. All authors revised the manuscript and provided their approval of the current version submitted for publication. All authors agree to be accountable for all aspects of the work, including its accuracy and integrity. Drs. Ray, Du, and Bujarski are co-guarantors of this review. To promote scientific transparency, study files are also made publicly available through github at <https://github.com/sbujarski>.

ADDITIONAL INFORMATION

Supplementary Information accompanies this paper at (<https://doi.org/10.1038/s41386-020-00913-3>).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

- Pickens R. Behavioral pharmacology: a brief history. *Adv Behav Pharmacol*. 1977;1:229–57.
- Miranda R Jr, Ray LA, O'Malley SS. The role of clinical (human) laboratory research in psychopharmacology. *APA handbook of psychopharmacology*. American Psychological Association; Washington, DC, 2019. p. 87–107.
- Yardley MM, Ray LA. Medications development for the treatment of alcohol use disorder: insights into the predictive value of animal and human laboratory models. *Addict Biol*. 2017;22:581–615.
- Litten RZ, Egli M, Heilig M, Cui C, Fertig JB, Ryan ML, et al. Medications development to treat alcohol dependence: a vision for the next decade. *Addict Biol*. 2012;17:513–27.
- Comer SD, Ashworth JB, Foltin RW, Johanson CE, Zacny JP, Walsh SL. The role of human drug self-administration procedures in the development of medications. *Drug Alcohol Depend*. 2008;96:1–15.
- Litten RZ, Falk DE, Ryan ML, Fertig J, Leggio L. Five priority areas for improving medications development for alcohol use disorder and promoting their routine use in clinical practice. *Alcohol Clin Exp Res*. 2020;44:23–35.
- Metz VE, Jones JD, Manubay J, Sullivan MA, Mogali S, Segoshi A, et al. Effects of ibudilast on the subjective, reinforcing, and analgesic effects of oxycodone in recently detoxified adults with opioid dependence. *Neuropsychopharmacology*. 2017;42:1825–32.
- Ray LA, Bujarski S, Courtney KE, Moallem NR, Lunny K, Roche D, et al. The effects of naltrexone on subjective response to methamphetamine in a clinical sample: a double-blind, placebo-controlled laboratory study. *Neuropsychopharmacology*. 2015;40:2347–56.
- Ray LA, Bujarski S, Shoptaw S, Roche DJ, Heinzerling K, Miotto K. Development of the neuroimmune modulator ibudilast for the treatment of alcoholism: a randomized, placebo-controlled, human laboratory trial. *Neuropsychopharmacology*. 2017;42:1776–88.
- Litten RZ, Falk D, Ryan M, Fertig J. Research opportunities for medications to treat alcohol dependence: addressing stakeholders' needs. *Alcohol Clin Exp Res*. 2014;38:27–32.
- Ray LA, Bujarski S, Roche DJO, Magill M. Overcoming the "valley of death" in medications development for alcohol use disorder. *Alcohol Clin Exp Res*. 2018;42:1612–22.
- Ray LA, Hutchison KE, Tartter M. Application of human laboratory models to pharmacotherapy development for alcohol dependence. *Curr Pharm Des*. 2010;16:2149–58.
- Mason BJ, Higley AE. A translational approach to novel medication development for protracted abstinence. *Curr Top Behav Neurosci*. 2013;13:647–70.
- Litten RZ, Wilford BB, Falk DE, Ryan ML, Fertig JB. Potential medications for the treatment of alcohol use disorder: An evaluation of clinical efficacy and safety. *Subst Abuse*. 2016;37:286–98.
- Maisel NC, 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*. 2013;108:275–93.
- Rösner S, Hackl-Herrwerth A, Leucht S, Vecchi S, Srisurapanont M, Soyka M. Opioid antagonists for alcohol dependence. *Cochrane Database Syst Rev*. 2010: CD001867.
- Ray LA, Green R, Roche DJO, Magill M, Bujarski S. Naltrexone effects on subjective responses to alcohol in the human laboratory: a systematic review and meta-analysis. *Addict Biol*. 2019;24:1138–52.
- Hendershot CS, Wardell JD, Samokhvalov AV, Rehm J. Effects of naltrexone on alcohol self-administration and craving: meta-analysis of human laboratory studies. *Addict Biol*. 2017;22:1515–27.
- Swift RM, Whelihan W, Kuznetsov O, Buongiorno G, Hsuing H. Naltrexone-induced alterations in human ethanol intoxication. *Am J Psychiatry*. 1994;151:1463–7.
- King AC, Volpicelli JR, Frazer A, O'Brien CP. Effect of naltrexone on subjective alcohol response in subjects at high and low risk for future alcohol dependence. *Psychopharmacol*. 1997;129:15–22.
- Volpicelli JR, Watson NT, King AC, Sherman CE, O'Brien CP. Effect of naltrexone on alcohol "high" in alcoholics. *Am J Psychiatry*. 1995;152:613–5.
- Mann K, Roos CR, Hoffmann S, Nakovics H, Lemenager T, Heinz A, et al. Precision Medicine in Alcohol Dependence: a Controlled Trial Testing Pharmacotherapy Response Among Reward and Relief Drinking Phenotypes. *Neuropsychopharmacology*. 2018;43:891–99.
- Miranda R, Jr, O'Malley SS, Treloar Padovano H, Wu R, Falk DE, Ryan ML, et al. Effects of alcohol cue reactivity on subsequent treatment outcomes among treatment-seeking individuals with alcohol use disorder: a multisite randomized, double-blind, placebo-controlled clinical trial of varenicline. *Alcohol Clin Exp Res*. 2020;44:1431–43.
- Ryan ML, Falk DE, Fertig JB, Rendenbach-Mueller B, Katz DA, Tracy KA, et al. A phase 2, double-blind, placebo-controlled randomized trial assessing the efficacy of ABT-436, a novel V1b receptor antagonist, for alcohol dependence. *Neuropsychopharmacology*. 2017;42:1012–23.
- Bujarski S, Hutchison KE, Prause N, Ray LA. Functional significance of subjective response to alcohol across levels of alcohol exposure. *Addict Biol*. 2017;22: 235–45.
- Ray LA, Bujarski S, Roche DJ. Subjective response to alcohol as a research domain criterion. *Alcohol Clin Exp Res*. 2016;40:6–17.
- Schuckit MA. Subjective responses to alcohol in sons of alcoholics and control subjects. *Arch Gen Psychiatry*. 1984;41:879–84.
- King AC, de Wit H, McNamara PJ, Cao D. Rewarding, stimulant, and sedative alcohol responses and relationship to future binge drinking. *Arch Gen Psychiatry*. 2011;68:389–99.
- Grodin EN, Montoya AK, Bujarski S, Ray LA. Modeling motivation for alcohol in humans using traditional and machine learning approaches. *Addict Biol*. 2020; e12949. <https://doi.org/10.1111/adb.12949>. Online ahead of print.
- Bujarski S, Ray LA. Experimental psychopathology paradigms for alcohol use disorders: Applications for translational research. *Behav Res Ther*. 2016;86:11–22.
- Ray LA, Roche DJ, Heinzerling K, Shoptaw S. Opportunities for the development of neuroimmune therapies in addiction. *Int Rev Neurobiol*. 2014;118:381–401.
- Ripley TL, Stephens DN. Critical thoughts on current rodent models for evaluating potential treatments of alcohol addiction and withdrawal. *Br J Pharmacol*. 2011;164:1335–56.

33. Bormann I. Digitizelt software, version 2.0. Braunschweig, Germany. 2012.
34. Rakap S, Rakap S, Evran D, Cig O. Comparative evaluation of the reliability and validity of three data extraction programs: UnGraph, GraphClick, and Digitizelt. *Computers Hum Behav.* 2016;55:159–66.
35. Bujarski S, Hutchison KE, Roche DJ, Ray LA. Factor structure of subjective responses to alcohol in light and heavy drinkers. *Alcohol Clin Exp Res.* 2015;39:1193–202.
36. Ray LA, MacKillop J, Leventhal A, Hutchison KE. Catching the alcohol buzz: an examination of the latent factor structure of subjective intoxication. *Alcohol Clin Exp Res.* 2009;33:2154–61.
37. Food and Drug Administration. Alcoholism: Developing Drugs for Treatment Guidance for Industry. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER). Retrieved from <http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm433618.pdf> 2015.
38. Magill M, Ray L, Kiluk B, Hoadley A, Bernstein M, Tonigan JS, et al. A meta-analysis of cognitive-behavioral therapy for alcohol or other drug use disorders: treatment efficacy by contrast condition. *J Consult Clin Psychol.* 2019;87:1093–105.
39. Ray LA, Meredith LR, Kiluk BD, Walthers J, Carroll KM, Magill M. Combined pharmacotherapy and cognitive behavioral therapy for adults with alcohol or substance use disorders: a systematic review and meta-analysis. *JAMA Netw Open.* 2020;3:e208279.
40. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw.* 2010;36:1–48.
41. Williamson J. Least-squares fitting of a straight line. *Can J Phys.* 1968;46:1845–47.
42. York D. Least-squares fitting of a straight line. *Can J Phys.* 1966;44:1079–86.
43. York D. Least squares fitting of a straight line with correlated errors. *Earth Planet Sci Lett.* 1968;5:320–24.
44. York D, Evensen NM, Martinez ML, De Basabe Delgado J. Unified equations for the slope, intercept, and standard errors of the best straight line. *Am J Phys.* 2004;72:367–75.
45. van Aert RC, Wicherts JM, van Assen MA. Conducting meta-analyses based on p values: reservations and recommendations for applying p-uniform and p-curve. *Perspectives on. Psychological Sci.* 2016;11:713–29.
46. Du H, Liu F, Wang L. A Bayesian “fill-in” method for correcting for publication bias in meta-analysis. *Psychological methods* 2017;22:799.
47. van Aert R. Puniform: meta-analysis methods correcting for publication bias. R Package Version. 2017;00:3.
48. Litten RZ, Falk DE, Ryan ML, Fertig JB. Discovery, development, and adoption of medications to treat alcohol use disorder: goals for the phases of medications development. *Alcohol Clin Exp Res.* 2016;40:1368–79.
49. Litten RZ, Ryan M, Falk D, Fertig J. Alcohol medications development: advantages and caveats of government/academia collaborating with the pharmaceutical industry. *Alcohol Clin Exp Res.* 2014;38:1196–9.
50. Garbutt JC, Kranzler HR, O'Malley SS, Gastfriend DR, Pettinati HM, Silverman BL, et al. Efficacy and tolerability of long-acting injectable naltrexone for alcohol dependence: a randomized controlled trial. *JAMA.* 2005;293:1617–25.
51. Jonas DE, Amick HR, Feltner C, Bobashev G, Thomas K, Wines R, et al. Pharmacotherapy for adults with alcohol use disorders in outpatient settings: a systematic review and meta-analysis. *JAMA.* 2014;311:1889–900.
52. Heilig M, Thorsell A, Sommer WH, Hansson AC, Ramchandani VA, George DT, et al. Translating the neuroscience of alcoholism into clinical treatments: From blocking the buzz to curing the blues. *Neurosci Biobehav Rev.* 2010;35:334–44.
53. Heilig M, Sommer WH, Spanagel R. The need for treatment responsive translational biomarkers in alcoholism research. *Curr Top Behav Neurosci.* 2016;28:151–71.
54. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Cochrane Database Syst Rev.* 2009;MR000006.
55. Easterbrook PJ, Gopalan R, Berlin J, Matthews DR. Publication bias in clinical research. *Lancet* 1991;337:867–72.
56. Egli M. Can experimental paradigms and animal models be used to discover clinically effective medications for alcoholism? *Addiction Biol* 2005;10:309–19.
57. Egli M. Advancing Pharmacotherapy Development from Preclinical Animal Studies. *Handb Exp Pharmacol.* 2018;248:537–78.
58. Jones JD, Comer SD. A review of human drug self-administration procedures. *Behav Pharmacol* 2013;24:384–95.