Title

Semi-parametric estimation of direct and indirect effects in the COMBINE (Combining Medications and Behavioral Interventions for Alcoholism) Study

Author

Subbaraman, Meenakshi Sabina

Publication Date

2012

Peer reviewed|Thesis/dissertation

Semi-parametric estimation of direct and indirect effects in the COMBINE (Combining Medications and Behavioral Interventions for Alcoholism) Study

By

Meenakshi Sabina Subbaraman

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Epidemiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:
Professor Jennifer Ahern, Chair
Professor Lee Ann Kaskutas
Professor Mark van der Laan
Professor David Presti
Spring 2012

Abstract

Semi-parametric estimation of direct and indirect effects in the COMBINE (Combining Medications and Behavioral Interventions for Alcoholism) Study

by

Meenakshi Sabina Subbaraman

Doctor of Philosophy in Epidemiology

University of California, Berkeley

Professor Jennifer Ahern, Chair

The investigators of the COMBINE (Combining Medications and Behavioral Interventions for Alcoholism) Study aimed to determine whether naltrexone, a drug alleged to reduce cravings for alcohol, combined with a behavioral intervention (CBI) assumed to change stress and coping behaviors, improves drinking outcomes more than either alone. After 16 weeks, only naltrexone alone and CBI alone significantly increased percent days abstinent (PDA) in models controlling for baseline PDA and site of treatment administration. Unexpectedly, the naltrexone + CBI combination did not offer any advantage over either naltrexone alone or CBI alone.

To understand moderating and mediating factors, and to help explain the Combination's lack of improvement over each monotherapy, controlled and natural direct effect analyses were performed using targeted maximum likelihood estimation (TMLE). TMLE offers several advantages over traditional direct effect analytic approaches such as double-robustness and allowance of treatment moderation by potential mediators. Cravings and stress were examined as theoretically informed mediators/moderators.

Controlled direct effect results show that naltrexone, CBI, and the Combination all work best when cravings are high, while none work when cravings are low. Similarly, naltrexone and the Combination work better when stress is high. Natural direct/indirect effect results show that all three treatments' effects are at least partially mediated by cravings, and that craving reduction explains 47-63% of treatment effects. Furthermore, naltrexone appears to affect cravings earlier while CBI works later.

Taken together, the set of results suggests the possibility of a threshold effect; if naltrexone reduces cravings early on and CBI is not effective when cravings are low, then the Combination's lack of improvement over either monotherapy should not be surprising.

DEDICATION

This dissertation is dedicated to the alcoholic who still suffers.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGMENTS

The analyses for this dissertation would have been impossible without Samuel Lendle, my biostatistics whiz , programmer extraordinaire, and good friend.

INTRODUCTION

Alcohol dependence pervades society. Alcohol dependence is responsible for numerous problems including traffic accidents; death to cirrhosis, liver disease, and other alcohol-related conditions; workdays lost; increased incidence of violent crimes, homicide and suicide; and adverse effects on lives of alcoholics and their families. However, less than 5% of alcohol dependent patients receive adequate treatment (Heinz, 2003).

Alcohol dependence is the uncontrollable compulsion to drink despite adverse consequences. Treatments for alcohol dependence vary across a broad spectrum. Myriad studies have shown that psychosocial treatments, behavioral treatments, 12-step based protocols, "holistic" programs, and prescription medications such as naltrexone, acamprosate, and antabuse all can decrease drinking and promote abstinence. Some psychiatrists and treatment professionals recommend a combination of treatment protocols, such as naltrexone with cognitive behavioral therapy, since combining treatments theoretically could offer an improvement over individual therapies.

This dissertation examines the COMBINE study, which was designed to assess whether combining a pharmaceutical treatment with a behavioral treatment would improve drinking outcomes more than either treatment alone. To provide background on the treatments applied in COMBINE, I will review literature regarding various alcohol dependence treatments, concentrating on the pharmaceuticals naltrexone and acamprosate, as well as cognitive behavioral therapy, a psychosocial intervention. I will then discuss the specific aims and hypotheses of my dissertation, as well as the methods used to assess my hypotheses. Finally, I will explain my findings and discuss the implications of my results.

I. SUBSTANTIVE BACKGROUND: EVIDENCE-BASED TREATMENTS FOR ALCOHOL DEPENDENCE

Pharmacologic treatments for alcohol dependence

Chronic alcohol consumption changes the brain's physiologic structure, as alcohol activates various neural networks; hence repeated exposure to alcohol modifies neural pathways as the brain attempts to restore homeostasis post-consumption (1). Pharmacologic treatments seek to interfere with the neural circuit regulating the desire to drink.

Antabuse

Introduced to the general public in 1948 antabuse (Disulfiram®), was the first available pharmacologic indicated for the treatment for alcohol dependence (2). The utility of antabuse, an organic sulfur compound, in the treatment of alcohol dependence was discovered by chance when rubber plant workers noticed they were unable to drink alcohol after work without becoming physically ill (3).

Antabuse is meant to be taken daily, and works by inducing severe physical discomfort if any amount of alcohol is ingested after taking the drug. Antabuse blocks the enzyme aldehyde dehydrogenase, which inhibits aldehyde oxidase of the liver and subsequently disrupts metabolism acetaldehyde, a byproduct of alcohol oxidation (4). This metabolic disruption causes acetaldehyde build-up, leading to physical symptoms including nausea, headache, heart palpitations, and trouble breathing. Theoretically, antabuse works by conditioning an adverse response to alcohol in the alcoholic. The goal of antabuse is to condition patients to avoid alcohol because it leads to unpleasant physical symptoms (5). Antabuse does not significantly affect craving for alcohol (4). The recommended dose is 250mg/day with a maximum of 500mg/day. Side effects of antabuse, besides its intended inducement of sickness if alcohol is ingested, include indigestion, dizziness, headache drowsiness, and tiredness (3).

One of the earliest reports on antabuse efficacy came from a hospital-based study of 112 treatment-seeking patients and stated that nine months post-treatment, younger patients, patients with slight reductions in social problems, and patients with none/minor psychoneurotic symptoms had "somewhat better" results than other patients; there were no other significant differences across patient subgroups (Kimber, 1950). In addition to the vague results, this trial did not include a control group and only consisted of 112 patients, limiting its power and generalizability.

In their 1971 review, Lundwall and Baekeland commented that most prior antabuse studies lacked methodological rigor (3), citing experimental design as the fundamental problem. Lundwall and Baekeland claimed that besides the absence of control groups in most trials of the time, many prior studies also focused too much on abstinence as an outcome (as opposed to reductions in drinking), followed patients for inadequate or arbitrary lengths of time, and consisted of overly heterogeneous patient populations. Most importantly, Lundwall and Baekeland reported that most studies did not perform statistical

tests, rendering results "in a strict sense, uninterpretable." After applying rigorous inclusion/exclusion criteria in evaluating extant antabuse trials, Lundwall and Baekeland concluded that only one, that of Wallerstein et al., contained scientifically valid results.

Wallerstein et al. (1957) compared results across four treatment groups followed for two years: antabuse, hypnotherapy, conditional reflex therapy and a control group. Wallerstein et al. examined a variety of outcomes in 178 patients including degree of abstinence, social adjustment, patients' subjective feelings of change, and psychiatric measures. "Overall improvement percentages," measured as changes in several dimensions including abstinence, job performance, and improvements in both intra- and inter-personal functioning were 53% for the antabuse group; 36% for hypnotherapy; 24% for conditioned reflex therapy; and 26% for the control group (3); however, no statistical tests were performed.

Noting that only one of the almost 100 prior antabuse trials was blinded, Fuller et al. conducted a blinded randomized control trial comparing 250 mg antabuse, 1mg antabuse, and vitamin (riboflavin) control conditions in 605 men followed for one year (Fuller, 1986). The authors found no significant differences in total abstinence, time to first drink, social stability, or employment (2). Fuller et al. performed blood and urine analyses to validate reported abstinence and found that although fewer patients in the 250 mg antabuse group screened positive for ethanol compared to the other treatment groups, the difference was not statistically significant. The authors claimed that previous results in favor of antabuse may have been induced by methodological flaws including prior use of incomparable control groups, poorly defined outcomes, small sample sizes, short follow-up periods, lack of blinding, and reliance on patient self-report. By implementing a large, blinded, randomized control trial with a relatively long follow-up period and blood/urine ethanol content validation analyses, Fuller et al. provided strong evidence that contradicted earlier reports of antabuse efficacy.

However, 1986 Fuller et al.'s results showed compliance to be a crucial factor in antabuse efficacy: 43% (50/116) of compliant participants maintained abstinence while only 8% (26/315) of the noncompliant men were abstinent after one year ($p<.001$). However, there were no significant differences in compliance across treatment groups: 23% of the 250 mg antabuse group, 17% of the 1 mg antabuse group, and 18% of control participants were compliant, respectively, and of these compliers, 38%, 50%, and 43% were abstinent, respectively. Thus although antabuse's lack of documented efficacy cannot be explained by differential rates in compliance across treatment conditions, perhaps the low rates of compliance within treatment groups limit power to detect antabuse effects.

Criticism of antabuse comes from Heilig et al. who stated that antabuse is outdated because it punishes drinking by inducing sickness if alcohol is consumed rather than targeting the dependence itself. According to Heilig, "for optimal efficacy, punishment must be applied severely and consistently. Taking this to a logical, but undesirable extreme, it would be more effective and safe to use a biosensor and an electric shock generator" (5). A 2008 German review by Mutschler et al. similarly concluded that antabuse should only taken in conjunction with "comprehensive therapy" under medical supervision. These conclusions corroborate early advice from Kimber (1950) who stated that antabuse should be prescribed as "only an adjuvant to be used in conjunction with other available methods. Antabuse given alone will achieve abstinence only as long as the

patient has the energy to take his tablets-hardly longer than the periods of abstinence so solemnly promised by any alcoholic" (Kimber, 1950).

A 1997 review of 38 antabuse studies found a general absence of methodological rigor, and reported that earlier studies had poorer validity than more recent (6). After reviewing studies published between 1967 and 1995, Hughes concluded that evidence for the efficacy of oral antabuse was ambiguous, and that most results in favor of antabuse regarded reduced consumption and number of drinking days (6). Hughes claimed that, "an effect in increasing the proportion of patients who achieve abstinence is surprisingly lacking." Like Mutschler, Hughes commented that antabuse should be used under close supervision as part of a more comprehensive treatment protocol.

Although results from antabuse trials appeared mixed, it remains the most widely used medication to treat alcohol dependence in the US (4). According to Lundwall and Baekeman, the initial excitement over antabuse stemmed from its novelty as the first pharmaceutical treatment, and potential fears induced by antabuse's alleged effects may have led to selection of more highly motivated patients, thereby artificially skewing early results in favor of antabuse efficacy while later, more methodologically rigorous studies yielded less favorable results.

Acamprosate

Amidst both the doubts over antabuse utility and advances in the understanding of neurochemistry, acamprosate (Campral®) was developed (7). FDA approved in 2004 but first available in France in 1989 (8), acamprosate is a structural analogue of the neurotransmitter gamma-aminobutyric acid (GABA) and an NMDA modulator, that acts as both an NMDA agonist and antagonist. NMDA, or N-methyl-o-aspartic acid, is an amino acid that behaves like the neurotransmitter glutamate on receptors in the brain (9). Glutamate, the most prevalent excitatory neurotransmitter in the brain, is thought to be involved in cognitive processes such as learning and memory formation (10). Antagonism of glutamate receptors is thought to reinstate proper neurotransmission that may have been dysregulated by chronic alcohol use (11). Acamprosate also inhibits GABA (γ-aminobutyric acid) function; GABA, the primary inhibitory neurotransmitter, is thought to regulate neuronal excitability. Interestingly, acamprosate is the only non-mu-opioid-receptor agonist/antagonist of medications indicated for treating alcohol dependence (1). These neurobiological properties along with clinical evidence suggest that acamprosate primarily works by decreasing cravings related to avoiding withdrawal symptoms (12).

NMDA modulators like acamprosate alleviate withdrawal and stabilize neurochemical changes resulting from chronic abuse and subsequent abstinence (13), as well as reduce response to abuse related stimuli. GABA receptor agonists are thought to increase amounts of available GABA in the brain, producing a relaxing, anti-anxiety state. According to 2008 FDA regulations, recommended dosage for acamprosate is usually 300 mg three times daily, but can be adjusted for bodyweight (14). Side effects of acamprosate include nausea, diarrhea, stomach pain, and confusion (9). Acamprosate does not enhance the toxic effects of alcohol, does not induce hepatotoxicity, and does not have abuse potential (4, 15).

Most clinical support for acamprosate efficacy comes from Europe. In one of the earliest published acamprosate trials Lhuintre and colleagues (1990) randomly and blindly

assigned 569 participants to receive either acamprosate or placebo, and examined effects on relapse using plasma gamma-glutamyl transpeptidase (GGT), the prevalent liver enzyme whose levels increase after alcohol ingestion, (Lamy et al., 1974) as a biomarker of alcohol intake. Three months post-treatment, those in the acamprosate group had significantly lower GGT plasma levels than those in the placebo group (1.4 ± 1.56 versus 2.0 ± 3.19 times normal, $p = 0.016$). Although non-significant, results for other indices of alcohol use, such as anxiety, trended in favor of acamprosate (9).

A 1995 trial led by Paille had similar results. Paille and colleagues performed a placebo-controlled, randomized double-blind study of two dosage levels of acamprosate in 538 alcohol-dependent patients recruited from 31 alcohol specialty centers. One hundred seventy-seven patients received placebo, 188 received 1.3 g/day of acamprosate (low dose group) and 173 received 2.0g/day (high dose group) for one year. With intention-to-treat analyses the investigators found evidence of a dose-response relationship: at all follow-ups, the highest abstinence rate was in the high dose acamprosate group, the lowest in the placebo group, and the low dose group in between. In terms of the primary outcome, the high dose acamprosate group had more days of continuous abstinence (153 ± 197) than either the low-dose group (135 ± 189) or the placebo group (102 ± 165; $p = 0.005$; Paille, 1995). At 6 months rates of abstinence were significantly higher in the high dose acamprosate group ($p < 0.02$) than either the low dose or placebo groups, with the lowest rates in the placebo group. The same results applied at 12 months, but were only borderline significant ($p = 0.096$).

In terms of secondary outcomes, acamprosate was associated with a prolonged period of initial abstinence (173 ± 126 days for placebo, 198 ± 133 days for low dose acamprosate and 223 ± 134 days for high dose acamprosate; $p = 0.032$); acamprosate patients were followed 75 days more on average than the patients on placebo ($p = 0.006$); acamprosate did not appear to significantly affect alcohol craving. Similarly to Lhuintre, Paille examined GGT levels along with other clinical indicators, such as blood alcohol level, for outcome verification, and found that that the percentage of subjects with normal GGT levels at 6 months was significantly higher in those who received acamprosate compared to placebo (42.8% versus 29.4%; $p = 0.011$). Results were similar at 12 months (35.3% versus 21.5%; $p = 0.015$), and for indicators of blood alcohol level.

A 2001 Spanish randomized, double-blind trial of acamprosate in 296 alcohol-dependent participants found that those treated with acamprosate had an average number of cumulative abstinent days that was 19 days longer than the placebo treatment group ($p = 0.0006$), and that the stable recovery duration, defined as the number of abstinent days between the last relapse into any drinking and the end of the trial, was 16 days longer on average among those treated with acamprosate ($p = 0.021$; (16)). Using survival analysis, the investigators also found that complete abstinence was achieved and maintained by 35% of acamprosate-treated patients and 26% of placebo-treated patients ($p = 0.068$). Similarly to the Lhuintre and Paille studies, mean serum GGT also significantly differed between acamprosate and placebo groups, favoring acamprosate efficacy. However, an important difference between the Gual study and previous studies is that in the Gual study acamprosate was prescribed from the start of alcohol withdrawal, whereas most other trials administered acamprosate after detoxification.

A 2000 meta-analysis by Kranzler and Van Kirk (17) examined all published trials of acamprosate, as well as naltrexone, and showed that both were modestly but

significantly related to treatment retention and/or drinking outcomes. Including only randomized, placebo-controlled trials using intention- to-treat samples, Kranzler and Van Kirk analyzed three outcomes: cumulative abstinent days (CAD), percentage of subjects reporting continuous abstinence throughout study period, and study retention rates. Kranzler and Van Kirk found that from 11 studies comprising 3204 participants, of those taking acamprosate, 11.4% (SE = .073) more were abstinent compared to those taking placebo. Furthermore, of 10 acamprosate studies comprising 3077 participants, those treated with acamprosate had 12.9% (SE = .088) more abstinent days and a 7.4% (SE = .071) higher treatment retention rate compared to those treated with placebo (Kranzler and Van Kirk, 2001). Hence all three outcomes were statistically significant, and subjects receiving acamprosate had abstinence rates 7% to 13% higher than those receiving placebo. The most robust effect of acamprosate appeared to be on CAD. However, the authors did detect significant heterogeneity across studies. Kranzler and Van Kirk found that sex was significantly related to CAD ($r^2 = 0.922$, $p = 0.001$): across the 10 acamprosate studies, those including fewer males appeared more efficacious in terms of CAD. Larger studies also appeared to show smaller effects of acamprosate on CAD. Because most acamprosate studies were conducted in different European countries, heterogeneity of results was not unexpected.

Two more recent meta-analyses also concluded that acamprosate was clinically efficacious (18, 19). Carmen et al. analyzed data from 33 acamprosate studies and reported that acamprosate was associated with a significant improvement in abstinence rate (Peto OR = 1.88, 95% CI: (1.57, 2.25), $p < 0.001$) and days of cumulative abstinence. Mann et al. reviewed 17 studies consisting of more than 4,000 patients and concluded that 6-month abstinence rates were significantly higher among those treated with acamprosate compared to placebo (pooled RR = 1.47, 95% CI: (1.29, 1.69), $p < 0.001$). The pooled difference in one-year abstinence rates between acamprosate and placebo was 13.3%, which yields 7.5 as the number-needed-to-treat (defined as the number of patients that need to be treated with acamprosate to prevent 1 additional bad outcome; (5)).

A review of acamprosate and opioid antagonists led by Soyka (2003) reported that 13/16 European placebo-controlled trials of acamprosate showed increased abstinence rates among those treated with acamprosate, and concluded that the number needed to treat was 8.15. Of the studies meeting validity standards such as randomization, blinding, pre-defined outcome criteria, and intention-to-treat analyses, rates of abstinence were significantly higher among those who received acamprosate compared to placebo. The Health Technology Board of Scotland (HTBS) meta-analysis of data from 17 studies, which included data from two unpublished negative acamprosate studies corroborated Soyka's conclusions, yielding an estimated NNT ("number needed to treat") of 13.3 (20). Soyka called the NNT of 13.3 "promising," since, for example, the NNT for the use of statins to prevent cardiovascular mortality ranges from 30 to 90 (14, 20).

The exact mechanisms of acamprosate are unknown; however, clinical evidence suggests that acamprosate decreases cravings for alcohol through conditioned withdrawal. Because acamprosate functions as a glutamate antagonist, one proposed explanation for acamprosate's efficacy is modulation of glutamatergic neural pathways (5), which counteracts the "hyperglutamatergic state" caused by alcohol withdrawal (21). Littleton (1995) reported that acamprosate reduced craving according to patient anecdotes. Furthermore, acamprosate may influence "conditioned withdrawal-induced craving" with

the neurochemical basis of acamprosate having similarities alcohol withdrawal; according to Littleton, these similarities may point to potential molecular mechanisms of action for acamprosate. Littleton goes on to say that the known effects of acamprosate appear to coincide with alcohol craving suppression, and that acamprosate might preferentially target "relief craving," or the urge to avoid negative withdrawal symptoms (5, 12).

Overall the literature supports acamprosate efficacy. Most positive results come from European studies, with effect sizes generally being small. Later trials appeared to have improved upon earlier trials by standardizing diagnostic criteria and behavioral treatment adjuncts (11). Furthermore, craving reduction may explain acamprosate effects.

Naltrexone

The development of naltrexone (Revia®, Depade®) coincided with that of acamprosate, perhaps due to the limited efficacy of antabuse and the rapid growth of the field of neuroscience. In 1994 naltrexone became the first FDA approved medication for alcohol dependence treatment in the fifty years after the development of antabuse (8, 22).

Alcohol consumption causes opioid peptide release in the brain, which subsequently leads to the release of dopamine in the mesolimbic neural pathway. The activation of dopamine in the mesolimbic pathway of the ventral tegmental area has been shown to be related to pleasure; procurement of sex and food, as well as psychoactive drugs, activate the release of neurochemicals such as dopamine. From an evolutionary standpoint, the pleasurable feelings instilled by the release of dopamine are thought to reinforce life-sustaining behaviors such as reproduction and eating (23).

Naltrexone is an opioid receptor antagonist that blocks opiates from attaching to opioid receptor sites in the brain. Naltrexone was originally intended for treating opiate addiction. However, a number of studies have shown that naltrexone also blocks dopamine from being released in regions of the brain such as the nucleus accumbens when alcohol is ingested (4). Although alcohol primarily facilitates $GABA_A$ modulation, it does bind to other neurotransmitter receptors, including opioid receptors. Whether alcohol directly binds to opioid receptors is unclear, as alcohol's antagonistic effects on opioid receptors may be a byproduct of $GABA_A$ modulation. In either case, naltrexone has been shown to reduce alcohol craving because, theoretically, blocking pleasure from alcohol subsequently blocks the desire to drink (23). Suggested administration of naltrexone is 50mg/day. Side effects include nausea, headache, and anxiety. Naltrexone does not enhance the toxic effects of alcohol and does not have abuse potential (15).

The earliest published results regarding naltrexone efficacy come from a pilot study in 30 alcohol-dependent veterans. Volpicelli et al (1990) reported that the 14 veterans treated with naltrexone and intensive psychotherapy were less likely to relapse and had fewer drinking days than the 16 treated with placebo (Volpicelli, 1990).

In 1992, the *Archives of General Psychiatry* published two landmark studies of naltrexone that replicated Volpicelli et al.'s preliminary findings (24, 25). FDA approval of naltrexone was based on these trials' results (22), which both demonstrated naltrexone efficacy in treating alcohol dependence. First, Volpicelli randomized 70 male patients to either 50 mg naltrexone or placebo, in conjunction with treatment and followed them for three months. Among those taking placebo, 54.3% relapsed, while among those taking naltrexone, 23% relapsed. An intriguing finding was that 95% (19/20) participants

receiving placebo fully relapsed after taking the first sip, whereas the same happened to only half (8/16) of the participants receiving naltrexone.

Results from O'Malley corroborated those from Volpicelli. O'Malley et al. randomized 97 men and women to combinations of naltrexone, placebo, and two manualized psychotherapies, either coping skills/relapse prevention therapy or supportive therapy without specific coping skills education. Results showed that after three months, patients taking naltrexone had higher continuous abstention rates, two-thirds the risk of relapse, half as many drinking days, and one-third the number of standard drinks, compared to placebo groups, with the naltrexone + supportive therapy condition obtaining the highest abstinence rates. Furthermore, those receiving naltrexone and either psychotherapy (i.e., coping skills/relapse prevention therapy or supportive therapy) had fewer drinking-related problems, and reduced severity of drinking-related problems compared to the placebo conditions. Among those who sampled alcohol, those in the naltrexone + coping skills therapy condition were significantly less likely to relapse compared to all other conditions.

The replication of results in different populations using different psychotherapies in conjunction with the medications bolsters the validity of Volpicelli et al.'s (1990) original findings: those in the 1992 Volpicelli trial were all male, mostly African American, required detoxification, and received intensive day treatment for one month followed by biweekly appointments for the next two months, while those in the 1992 O'Malley trial included women, were mostly white, and received either weekly coping skills/relapse prevention therapy or supportive therapy (24-26).

A 1997 follow-up study by Volpicelli replicated the group's earlier findings. Because the 1992 O'Malley and Volpicelli studies had found that the naltrexone treatment effect was stronger for compliant participants, Volpicelli designed a pragmatic trial to study naltrexone effectiveness in conjunction with psychotherapy in a naturalistic setting. Volpicelli et al. randomized 97 participants to receive either naltrexone or placebo, both in addition to individual counseling. Results again showed that those receiving naltrexone had moderately better drinking outcomes than those receiving placebo. Similarly to the 1992 findings, results also indicated that compliant participants receiving naltrexone had the best outcomes. Intent-to-treat analyses showed that participants receiving placebo who completed treatment drank on significantly more study days (12.7%) compared to those receiving naltrexone who completed treatment (5.40%, $p = .03$). Among compliant participants, only 4/28 (14%) of those receiving naltrexone relapsed compared to 12/23 (52%) of the compliant participants receiving placebo. There were no significant differences in returns to drinking or sampling alcohol between the naltrexone and placebo conditions among less compliant participants.

The Health Technology Board for Scotland meta-analysis (2003) of naltrexone studies included 12 positive and 5 negative studies from both Europe and the US and concluded that the NNT to prevent relapse, defined as consumption of more than 5 drinks in one day, was 12.4 (20).

Despite the overwhelming evidence in favor of naltrexone, some studies' results conflict with the mass of data supporting naltrexone's efficacy. First, Krystal et al. (27) blindly randomized 627 severely dependent veterans (mostly male) to receive 3 months of naltrexone, 12 months of naltrexone, or placebo and assessed time to first day of heavy drinking (six or more drinks for men and four or more drinks for women) as the primary

outcome in the first 13 weeks of the trial. All patients received individual 12-step facilitation counseling and encouragement to attend Alcoholics Anonymous. With the objectives of determining whether short-term use of naltrexone decreased time to relapse compared to placebo, or whether long-term use of naltrexone decreased number of drinking days and/or number of drinks per drinking day compared to either short-term naltrexone use or placebo, Krystal et al. found no significant differences in any drinking outcomes across the three groups. After the first 13 weeks, there were no significant differences in time to relapse between the two naltrexone groups and the placebo group (CI for difference in days till relapse: -3.0, 22.8). Furthermore, at the one-year follow-up there were no significant differences in number of drinking days or number of drinks per drinking day across the three groups. Krystal et al. concluded that they did not have evidence of naltrexone (combined with a psychosocial program) efficacy in males with chronic, severe alcohol dependence, but warned that a different dose of naltrexone or its use in combination with another therapy may prove beneficial. Although the study was larger and had longer follow-up periods than previous studies, the patient population consisted of older, male, chronic alcoholic veterans, limiting the generalizability of these findings (27).

In the Women's Naltrexone Trial, O'Malley et al. (28) randomized 103 women with current alcohol dependence (some with comorbid eating disorder pathology) to either naltrexone or placebo in addition to group therapy for 12 weeks, and assessed time to first day of drinking and time to first day of heavy drinking, defined as consuming four or more drinks in one occasion. Although O'Malley et al. found no significant differences between the naltrexone and placebo conditions in time to the first drinking day ($p = 0.51$), time to first heavy drinking day ($p = 0.88$), or the percentage of subjects who continued to meet criteria for alcohol dependence during treatment ($p = 0.80$), naltrexone was associated with a significant delay in the time to the second ($p = 0.02$) and third drinking days ($p = 0.04$). Because all women received some cognitive behavioral therapy, and because overall the percentage of days abstinent more than doubled while the number of drinks per drinking day decreased from 7.12 drinks/day to 1.83 drinks/day throughout the course of treatment, the authors speculated that the behavioral therapy may have had some overall effect, obscuring any differences between medication groups. Among the women with eating disorder pathologies, there were no significant differences between naltrexone and placebo on eating disorder measures (28).

However, a semiparametric trajectories re-analysis of the Krystal et al. (2001) and O'Malley (2007) data estimated three distinct trajectories of daily drinking: abstainers, sporadic drinkers and consistent drinkers (29). Gueorguieva et al. concluded that in the Krystal et al. (2001) study, those receiving naltrexone had twice the odds of following the abstainer trajectory over the consistent drinker trajectory but were not significantly more likely to follow the abstainer trajectory over the sporadic drinker trajectory (29). Gueorguieva also found statistically significant effects of naltrexone on increasing the likelihood of abstinence and decreasing the likelihood of heavy drinking in Krystal et al.'s sample. Re-analyses of the Women's Naltrexone Trial resulted similarly, but non-significantly. Gueorguieva et al. hypothesized that the original studies did not uncover a significant naltrexone effect because of the overall high abstinence rates and use of summary consumption measures. The authors also speculated as to why they discovered a significant result in men but not women, and concluded that the discrepancy could be due

to the larger VA sample Krystal et al. employed. Furthermore, some women included in the O'Malley study had comorbid eating disorders. Gueorguieva et al. also stated that treatment compliance appeared to decrease odds of drinking regardless of treatment, as controlling for daily adherence strengthened the naltrexone effects; this coincides with suggestions from Volpicelli (1992, 1997). The authors recommended trajectory-based methods for analyzing clinical trial data because of its value in estimating and accounting for study sample heterogeneity and identifying patient subgroups with similar response patterns for which treatment is effective.

Heilig et al. concluded that meta-analyses of available trials, even when the Krystal et al. negative findings are included in the analysis, unambiguously support naltrexone efficacy. Results from numerous meta-analyses and reviews support this conclusion (4, 19, 22) (30). The literature overwhelmingly favors naltrexone efficacy, especially among compliant patients and those with high baseline levels of craving.

Behavioral therapies for alcohol dependence

The goal of behavioral treatments, besides abstinence, is to teach participants to cope with stress or develop healthier leisure-time activities. Behavioral treatments revolve around the idea that drinking behaviors are positively reinforced in both social and physiologic ways (31). Although the mediating mechanisms of behavioral treatments have not been well established, numerous studies have shown that behavioral therapy is effective long-term (32, 33). According to results from a 2007 review, no one behavioral therapy appears superior to others (34). Furthermore, although controversial, results from the large, multi-site Project MATCH appear to reflect comparable efficacy across motivational interviewing, twelve-step facilitation, and cognitive behavioral approaches (35). Noting the strengths of each MATCH therapy, COMBINE study investigators developed the Combined Behavioral Intervention (CBI) by integrating elements of the three behavioral studies implemented in Project MATCH. In this section, I briefly summarize Project MATCH, describe these three common behavioral approaches in the context of the MATCH studies, and conclude with a description of the COMBINE study CBI.

Project MATCH overview

Project MATCH (Matching Alcohol Treatments to Client Heterogeneity) was the largest trial of psychotherapies for alcohol dependence to date, *with the primary goal of determining whether "matching" alcohol-dependent patients based on personality and other attributes to various treatment modalities would result in increased treatment efficacy* (35, 36).

Investigators followed 952 outpatient and 774 aftercare clients (N = 1726) for three years after randomization to 12 weeks of individual sessions of motivation interviewing, twelve-step facilitation, or cognitive behavioral therapy. MATCH investigators utilized these particular treatments because of established clinical effectiveness, distinctiveness from one another, and ease of application in existing treatment centers. In the next sections, I briefly describe Project MATCH results and treatment conditions (motivational interviewing, twelve-step facilitation, and cognitive-

behavioral therapy), because the COMBINE study developed its combined behavioral intervention (CBI) from the Project MATCH treatments.

Project MATCH investigators examined ten client attributes as potential treatment modality modifiers: severity of alcohol involvement; cognitive impairment; client conceptual level; gender; meaning seeking; motivational readiness to change; psychiatric severity; social support for drinking versus abstinence; sociopathy; and typology (35). Sixteen matching hypotheses were put forth; for example, clients with high motivation and those with much network support for drinking were expected to benefit more from cognitive behavioral therapy compared to motivational interviewing, while those high on meaning seeking were expected to benefit more from twelve-step facilitation than either motivational interviewing or cognitive behavioral therapy.

Project MATCH results shook the field. Contrary to primary hypotheses, the only patient attribute that appeared to "match" to treatment efficacy after one year was psychiatric severity: outpatients with low psychiatric severity had higher percent days abstinent if treated with twelve-step facilitation (TSF) compared to cognitive behavioral therapy. As psychopathology worsened, the TSF advantage disappeared. TSF benefits also disappeared three years posttreatment. Of the 21 patient attributes studied in relation to outcomes, readiness to change and self-efficacy were the strongest predictors of long-term drinking (37).

Overall, TSF patients seemed to have slightly higher percentage days abstinent than patients from the other two treatment conditions. However, no difference was found for drinks/drinking day. In the aftercare arm, those high on meaning seeking appeared more likely to have more abstinent days in the second half of the one-year posttreatment period if treated with TSF. Although a modest finding, this result is theoretically sound as the twelve-step emphasis on spirituality was expected to resonate in those with high meaning seeking. However, investigators concluded that no matching hypotheses were supported among aftercare patients.

Three-year posttreatment results showed that outpatients with high anger levels appeared to do better in MET than CBT or TSF: patients in the highest third of the anger construct had 76% abstinent days if treated with MET compared to 66% for those treated in CBT or TSF (37). This association was also present at one year, making it the most consistent "match" (38). Aftercare patients with high alcohol dependence fared better in TSF than CBT at the three-year follow-up as well; this finding may be partially explained by the relatively increased focus of TSF on abstinence as a goal compared to MET and CBT (39). Furthermore, outpatients with higher social support for drinking fared better in TSF (83% abstinent) than MET (66% abstinent) three years after treatment; this difference of 17% was the largest found (40), and partially explained by increased AA meeting attendance observed among those treated in TSF compared to the other treatments (37).

TSF also appeared to most benefit the sample overall: 36% of TSF patients were abstinent three years posttreatment compared to 27% of MET and 24% of CBT patients (37). Importantly, the authors also stated that they could not claim treatment efficacy for any of the three conditions due to the absence of a control group (41). However, posttreatment drinking levels did show an overall change from 25 drinking days/month to 6 drinking days/month as well as a five-fold reduction in volume across the three treatment groups (40).

Notably, *causal chain analyses showed that all failed matches (i.e., lack of support for hypothesized moderation/increased treatment efficacy by subgroups based on personality and other attributes to each treatment condition) were due to the dysfunctionality of hypothesized causal mechanisms* (40). In other words, subgroups expected to benefit most from particular MATCH treatment conditions did not appear to gain any advantages from *a priori* hypothesized "matches," and these unexpected results may be due to breakdowns in the theoretical causal pathways. MATCH treatments may not be affecting causal intermediates as they are theoretically expected to do. Interestingly, even when matching hypotheses were supported, mechanisms were supported only half the time; hypothesized causal mechanisms were shown to be at work in the case of those with high alcohol dependence or support for drinking (both did better in TSF), but not for those with low psychiatric severity (better in TSF) or anger (better in MET).

The disappointing results from Project MATCH have incited much discussion. A secondary analysis by Cutler and Fishbain (2005) showed that although treatments were comparable in terms of efficacy, none of Project MATCH's treatments were particularly effective; any improvements in drinking outcomes were due to trial enrollment and client selection effects. First, treatment attendance and outcomes were only weakly correlated, and a median of only 3% variance in outcomes could be explained by actual treatment attendance. Furthermore, drinking improvements occurred in the initial week post-baseline, but this improvement did not change across time; the initial improvement was maintained for 12 weeks of schedule treatment, whether the patients attended all sessions, only one session, or no treatment sessions at all. For example, percent days abstinent increased by more than 60% in the first week of treatment, but only 4% more in the 11 subsequent weeks (42).

Cutler and Fishbain argued that if treatment actually led to better outcomes, then the percent days abstinent should have improved over the full course of treatment as opposed to solely during the first week. Furthermore, those who had not completed any treatment sessions had similar percent days abstinent outcomes compared to those who completed all sessions. In terms of selection effects, Cutler and Fishbain contended that patients who have reduced consumption are more likely to finish treatment, while those who continue to drink are more likely to have dropped out. In other words, the initial improvements seen across all treatment conditions may be explained by selection bias; those with higher baseline levels of drinking may have attempted to attend the first treatment sessions because of their relatively severe drinking pattern but then dropped out due to continued drinking. Cutler and Fishbain concluded that ineffective treatments thus explain the surprising MATCH results, and suggested that not only were there no successful treatment matches, but that all treatments were equally effective (i.e., ineffective). In essence, Cutler and Fishbain contended that any improvements seen in the Project MATCH data are simply due to selection biases and cannot be attributed to the actual MATCH treatments (42).

Motivational interviewing

Motivational interviewing, one of the components of/treatments in MATCH, stems from motivational psychology and involves counseling sessions in which the therapist guides the patient's train of thought and provides structured feedback such that motivations for abstinence arise internally and become clear to the patient (42). Through interviewing techniques, the therapist is able to focus the patients' thoughts on reasons for why sobriety might be a good choice and guide the patient to operationalize motivational strategies to exploit his or her own resources. By understanding and focusing on fundamental reasons for an abstinent lifestyle, the patient is able to recognize the importance of sobriety and is hence expected to incorporate these ideals into his/her way of life by committing to change, accepting responsibility, and mobilizing individual resources (35, 42).

Project MATCH results showing that angry patients fared better in MET, in conjunction with the non-confrontational nature of MET, may indicate self-change processes as a potential MET mechanism of action (43). Mediation analyses indicated that lower levels of therapist directiveness in MET explained its advantage over CBT for high-anger patients (44). However, MATCH investigators found no evidence supporting causal mechanisms of MET, such as therapy session attendance and therapist confrontation (as measured from he Therapy Process Rating Scale (TPRS)), when investigating reasons for the slight increase in abstinent days observed in the last month of the first year of follow-up among low-motivation outpatients treated in MET relative to CBT (45).

Twelve-step facilitation

Twelve- step facilitation (TSF) approaches to treating alcohol dependence attempt to increase twelve-step involvement and retention by easing patients into the twelve-step lifestyle. Although TSF interventions vary in modality, Project MATCH TSF particularly focuses on the first three of the twelve steps of Alcoholics Anonymous (AA). MATCH TSF assumes that alcoholism is a disease, both medically and spiritually, and aims to convince patients they are powerless over alcohol and must abstain, and that AA meetings can help them do this.

Overall (i.e., without taking hypothesized "matched" subgroups into account), those in Project MATCH TSF had a 24% abstinence rate one year after treatment (compared to 14% in CBT and 15% in MET; (40)). However, MATCH TSF clients with social networks supportive of drinking also maintained abstinence more successfully than their counterparts treated with CBT or MET (46). Secondary causal chain analyses that attempted to explain these effects showed that TSF led to greater overall AA participation than CBT or MET: 75% vs. 35% and 38%, respectively. AA participation also significantly predicted more abstinent days. Mediation analyses examining the stronger effect of TSF observed among patients with social networks supportive of drinking confirmed that TSF intervention effects are best explained by increased 12-step meeting attendance or involvement. Enhanced AA involvement may thus explain the relatively higher success rates observed in the TSF condition, and implies that TSF may be the optimal treatment for those with networks supportive of drinking (46). Finney notes that the relatively comparable efficacy across therapies observed in Project MATCH holds clinical relevance in that treatment rooted in the 12-step philosophy can benefit patients as much, if not more, than CBT (47).

Cognitive behavioral therapy

The foundation of CBT rests on the idea that if patients change the way they think, they can change the way they behave (31). CBT assumes that drinking is a function of life problems and that by identifying and evaluating high-risk situations and detrimental thoughts, patients can recognize situations in which they might feel compelled to drink. The initial identification of negative thoughts is followed by an analysis of why the thought might be wrong and what the patient can do to correct it. The patient is subsequently encouraged to practice thought evaluation with the goals being to develop skills to identify high-risk situations for drinking and to apply new coping behaviors to deal with stressors without resorting to alcohol use.

Results from Project MATCH showed that overall (i.e., without accounting for hypothesized matches), TSF and CBT groups both had better outcomes than the MET group, but that TSF and CBT appeared equivalent to each other: at the 12-week follow-up 41% of those in either TSF or CBT were abstinent or drinking moderately while only 28% of those receiving MET could say the same. However, differences in abstinence outcomes across conditions changed over time; at the one- and three-year follow-ups, slightly more TSF patients were abstinent than CBT or MET.

As none of the matching hypotheses regarding CBT were supported by MATCH results (35), some suggest that CBT did not operate via its theoretical mechanisms (48) in that CBT did not increase drink refusal coping skills as intended (46). Although MATCH results showed that patients who reported greater drink refusal skills also reported better drinking outcomes, CBT clients did not appear to have used drink refusal skills more than those treated with TSF or MET (49). Other work by Longabaugh also suggests that enhanced coping skills do not mediate CBT effectiveness, and that the mechanisms that explain why CBT works remain unknown; in their review of ten studies examining CBT mediators, not one provided evidence supporting theoretical mechanisms (50).

COMBINE Combined Behavioral Intervention (CBI)

The COMBINE investigators built their behavioral intervention upon Project MATCH foundations. In developing CBI, COMBINE investigators merged "what [they] believed to be the best components of the MATCH treatments into a new, single, state-of-the-art treatment" (51). Investigators therefore combined what they considered the most crucial ingredients of each of the three MATCH therapies: MET styles and clinical techniques, CBT coping strategies, and encouragement to attend AA and other self-help groups. An added component allowed inclusion of patients' significant other(s). CBI thus consisted of motivational strategies to target patient readiness to change; skills training to improve issues like communication and life problems; and promotion of social support, especially that found in 12-step and other support groups (52).

The Combine Behavioral Intervention (CBI) included four phases. Phase 1 focused on increasing motivation. Phase 2 involved creating a treatment plan, discussing the importance of abstinence, and encouraging support group involvement. Phase 3 involved verbal and signed agreement to the treatment plan, as well as actually executing the plan. Phase 3 continued until all treatment modules that had been selected by the patient are

completed. Patients could select from the following modules based on their needs: assertiveness skills; communication skills; coping with craving and urges (chosen most often, by 57% of patients); drink refusal and social pressure; job finding (chosen least often, by 3% of patients); mood management; mutual support group facilitation; social and recreational counseling; and social support for sobriety. Phase 4 consisted of maintenance check-ups and progress review (53).

CBI therefore fell under the following structure. Patients first explored their problems and received motivational feedback from the therapist. Problems were systematically identified and assessed, antecedents and consequences of drinking were analyzed, and the patient and therapist then developed goals and plan for attaining those goals. Importantly, participants understood that the goal of treatment was abstinence, as opposed to controlled drinking. Patients created treatment plans based on the series of modules designed to address alcohol problems. The number of modules as well as timing of treatment sessions was discussed and carried out over the 16-week treatment period (54).

COMBINE investigators considered several pressing issues when developing CBI. First, they required treatment protocol that was both standardized and able to be delivered consistently, as well as flexible enough to meet individual patient needs. In this effort to balance standard vs. individualized treatment, investigators provided therapists with decision trees and clients with choice options. The decision tree permitted replication of the treatment procedure with similar patients, if treatment were shown to be effective. Allowing patients to participate in planning his/her treatment goals also increased motivation, a primary goal of MET. Hence treatments were delivered according to patient symptoms, stage of change, and preference, and the number of sessions and order of module delivery varied according to patient needs (54).

Matters that may have arisen despite patient planning were also addressed with CBI modules. For example, COMBINE therapists were able to follow a standardized protocol if patients resumed drinking, or were not taking medication properly. In these situations, the therapist would consult a "pull-out procedure" relevant to the issue. The standardization of these pull-outs ensured replication in clinical practice. Notably, the most often used pull-out (used for almost 25% of CBI patients) was that which addressed resumed drinking (54).

THE COMBINE STUDY

Rationale for combining treatments

Prior evidence supports examining acamprosate and naltrexone in combination, as both have shown to be clinically safe, moderately effective, and highly tolerable by patients (55, 56). Furthermore, alcohol activates various neurotransmitter systems, which suggests that effective therapy should target multiple neural receptors (57), and each medication works via distinct mechanisms. Naltrexone is an opioid receptor antagonist. Because the opioid receptors are known to mediate the rewarding effects of alcohol, naltrexone is alleged to attenuate the desire or craving for alcohol rewards (4). Although less understood, acamprosate is thought to modulate NMDA dysregulation, normalizing

glutamatergic excitation that occurs in withdrawal and early abstinence (4, 58). Acamprosate most likely reduces the desire or craving for reducing tension, which highlights the mechanistic distinction between naltrexone and acamprosate: *the primary difference between the alleged mechanisms of action appears to be alcohol craving reduction for naltrexone vs. anxiety reduction/withdrawal symptom alleviation for acamprosate*. Therefore the combination of naltrexone and acamprosate may synergistically affect craving by modulating both the neurotransmitters related to drinking triggers as well as those related to the conditioned responses to drink even after long periods of abstinence (56).

In addition to facilitating abstinence, the combination may help patients avoid relapse after ''slips" or the first sampling of alcohol after a period of abstinence. While naltrexone helps obtain initial drinking cessation, acamprosate helps maintain abstinence (58). Acamprosate has also shown to increase blood levels of naltrexone, subsequently enhancing naltrexone's neurochemical effects and potentially causing biological synergy (11, 58).

Ross (2009) also recommended administration of different therapies for particular stages of abstinence because medications could theoretically be combined to counteract both positive reinforcement (i.e., naltrexone) and negative reinforcement (i.e., acamprosate) from alcohol (57). Kreek et al. (2002) offered similar advice, stating that there are several time points for optimal pharmacologic treatment administration: during active drinking; during withdrawal or detoxification; during abstinence, as a method of relapse prevention. Kreek et al. proposed that combining pharmacologic agents or administering different medications at each time point can optimally target relevant issues at the appropriate times (e.g., negative reinforcement during withdrawal).

A 2003 German study was among the first to explore the acamprosate and naltrexone combination's efficacy. Kiefer et al. (2003) compared the combination to each drug monotherapy as well as a placebo condition. The investigators randomized 160 recently detoxified patients to blindly receive acamprosate + naltrexone, acamprosate alone, naltrexone alone, or placebo over the course of 12 weeks. All patients attended weekly group psychotherapy aimed at teaching participants to identify and cope with high-risk drinking situations. The investigators assessed time to first drink, time to relapse, and cumulative abstinence time using intention-to-treat survival analyses and found that for nonrelapse rates, both the naltrexone and acamprosate monotherapy conditions had superior outcomes compared to placebo ($p = .02$, $p = .05$, respectively), as did the acamprosate + naltrexone condition ($p = .008$). Although there were no significant differences in nonrelapse between the naltrexone and acamprosate monotherapies, the combination was more effective than acamprosate alone ($p = .04$), but not naltrexone alone. Treatment groups also significantly differed in time to first alcohol intake: the naltrexone and acamprosate monotherapies, as well as the combination, were again more effective than placebo ($p = .03$, $p = .04$, $p = .002$) respectively). There were no significant differences in time to first drink between the naltrexone and acamprosate monotherapies, but the combination was more effective than acamprosate alone ($p = .04$), but not naltrexone alone (59).

According to Heilig (2006), medications for treatment of alcohol dependence are best used in conjunction with behavioral treatments; Heilig based this statement on results from a meta-analysis that showed that in the short-term, naltrexone was not more effective

when combined with an intensive behavioral treatment compared to a less involved behavioral treatment; however, over a longer period of time, the intensive behavioral treatment did appear to enhance naltrexone's efficacy by increasing time to first drink and decreasing cravings (5).

According to Weiss and Kueppenbender, any medication administration inherently involves psychosocial interactions, and the psychosocial context of pharmacologic treatment may easily influence the medication's efficacy (60). The authors therefore urged exploration and identification of optimal psychosocial/psychopharmacologic combinations. In their review of combination therapies, Weiss and Kueppenbender concluded that prior studies suggest that naltrexone and cognitive behavioral therapy may interact positively, as naltrexone reduces the risk of a "slip" developing into an episode of heavy drinking (61), and cognitive behavioral therapy teaches skills for coping with slips and avoiding relapse. The authors also concluded that extant evidence shows that acamprosate appears equally effective across psychosocial treatment combinations, and that medical management may suffice for optimal acamprosate efficacy. If acamprosate, naltrexone and behavioral therapy affect abstinence via different pathways, perhaps some combination could target more known causal mechanisms than any treatment could individually.

Consequently, the primary goal of the COMBINE study was to examine whether combining medication and behavioral treatment would improve drinking outcomes compared to either medication or a cognitive behavioral intervention (CBI) alone.

COMBINE *study details*

COMBINE is the largest study of pharmacotherapy for alcoholism in the United States to date (62). COMBINE investigators randomized 1,383 newly abstinent participants to nine groups. Participants were predominantly white, male, middle-aged, married/in a stable relationship, relatively well educated, and employed (53). Enrollment criteria included: minimum average weekly alcohol consumption of 14 drinks for women or 21 drinks for men within a 30-day period in the 90 days prior to baseline; at least two heavy drinking days in the 90 days prior to baseline; and zero breath alcohol level before signing consent. Exclusion criteria included: meeting DSM-IV criteria for any psychiatric disorder requiring medication; opiate dependence/abuse in six months prior to baseline; current drug dependence (besides marijuana or nicotine); more than seven days inpatient treatment for substance use disorders in 30 days before randomization; planned continuation in pre-occurring alcoholism treatment during treatment phase of study; or abstinence from alcohol for more than 21 consecutive days prior to randomization (53).

Over a 16-week period, patients received either active/placebo naltrexone, or active/placebo acamprosate (total of 4 medication conditions: placebo, naltrexone, acamprosate, and naltrexone + acamprosate) either with or without the combined behavioral intervention (4 medication regimens x 2 CBI regimens = 8 groups). The ninth and final group underwent CBI only, without placebo or active medication.

Medical Management

	Placebo	Acamprosate
Placebo	1	2
Naltrexone	3	4

Medical Management + Psychotherapy

	Placebo	Acamprosate	No pills
Placebo	5	6	–
Naltrexone	7	8	–
No pills	–	–	9

FIGURE 1. COMBINE treatment combinations (Cell 9 patients received only CBI, no MM)

*Figure taken from Mattson and Litten, Combining Treatments for Alcoholism: Why and How? J. Stud. Alcohol, Supplement No. 15: 8-16, 2005

The active naltrexone groups received 100 mg doses, the active acamprosate groups received 3 mg doses, and those in CBI groups received up to 20 therapy sessions. The CBI only (no pills) group had access to 4 sessions with a health care professional who provided health care advice, while all participants treated with pills (either active medication or placebo) received Medical Management (MM), a low-intensity protocol intended for use in routine clinical practice. CBI, on the other hand, reflects treatment delivered in addiction specialty programs. During the medical management sessions, a health care professional provided information about the medications and adherence, and recommended support group attendance as well as abstinence from alcohol and drug use . One of COMBINE's goals was to determine whether the moderate-intensity CBI improved patient drinking beyond MM (53). Investigators further hypothesized that CBI would enhance pharmacologic effectiveness.

Treatment adherence had a median of 96.4% across all pill-taking groups (including placebo) and the attrition rate was 6%. Participants were interviewed 9 times during the 16-week treatment period and at 26, 52, and 68 weeks after the treatment period ended. Interviews assessed physical, physiologic, and laboratory measures; expectancies of treatment; behavioral measures; alcohol and drug involvement; craving; motivation; and psychological and psychiatric measures. The two primary outcomes were time to first heavy drinking day and percent days abstinent in the 16-week treatment period.

For analysis, COMBINE investigators used 8 of the 9 treatment combinations to create a 2x2x2 design that allows assessment of the effects of the individual therapies as well as comparisons of each combination. The investigators performed two intent-to-treat analyses of the 8 groups who received medication (either active or placebo) using percent days abstinent (PDA) and time to first heavy drinking day as co-primary outcomes. The investigators implemented a mixed-effects general linear model to assess treatment effects

on PDA, and a proportional hazards model to examine time to first heavy drinking day (53).

COMBINE *results*

Abstinence rates in the aggregate COMBINE sample were high, in fact higher than previously reported rates from similar trials (63): results showed that PDA increased from 25% to 73% between baseline and the study's end. In terms of treatment effects, the investigators found that compared to the placebo-only group (i.e., inactive pill without CBI), those taking only naltrexone had the highest odds of good clinical outcomes after one year; those who received either CBI + placebo or the combination of naltrexone + CBI had the next highest odds; and the remaining groups has the lowest odds of good clinical outcomes. There were no meaningful differences across the naltrexone, CBI, and naltrexone + CBI arms. The investigators categorized "good composite clinical outcomes at end of treatment" as abstinence, moderate drinking without problems, or neither (i.e., drinking with problems).

Although naltrexone alone and placebo + CBI significantly increased both PDA and time to first heavy drinking day, the Combination was no more efficacious than either alone. Surprisingly, no treatment combination including acamprosate improved outcomes significantly. Odds ratios for good composite clinical outcomes at end of treatment compared to the placebo-only group were 1.82 (95% CI: 1.26, 2.65) for placebo + CBI; 1.93 (95% CI: 1.33, 2.80) for naltrexone + CBI; and 2.16 (95% CI: 1.46, 3.20) for naltrexone alone (53).

A 2008 analysis of the COMBINE long-term outcomes across the one-year post-treatment follow-up period assessed changes in treatment effects over time (63). Donovan et al. examined PDA, time to relapse to heavy drinking, and "good clinical outcomes" during the 52 weeks following the end of the 16-week active-treatment period, and found a significant decrease in PDA across time, regardless of treatment group (PDA at week 26 > PDA at week 52, $p < .001$; PDA at week 26 > PDA at week 68, $p < .001$; PDA at weeks 52 and 68 did not differ, $p > .40$). There were no significant time*treatment interaction effects for PDA, although naltrexone alone or in combination with acamprosate and CBI, and CBI with double placebo all had higher levels of PDA than the double placebo with no CBI group.

Donovan et al.'s (2008) results also showed that after the initial 16-week treatment period, those who had been treated with naltrexone had significantly lower relapse rates than those who had been treated with placebo (hazard ratio = .74, 95% CI: .56-.97, $p = .03$). Furthermore, a significant main effect was found for CBI vs. no CBI in terms of "good clinical responders" (i.e., participants who were either abstinent or drinking moderately without problems): those who had been treated with CBI were almost 20% more likely to have a good clinical outcome through week 68 than those who had not received CBI ($p = .03$).

Thus Donovan et al. (2008) reported that CBI and/or naltrexone had better efficacy (compared to double placebo) in terms of drinking outcomes past the 16-week active-treatment phase, noting a small but significant advantage for CBI. The authors also stated that the positive effect of naltrexone on preventing relapse remained significant through week 52, implying a sustained benefit of the medication past 16 weeks. Furthermore, the

extended effects of CBI on good clinical outcomes may have stemmed from the fact that effects of some psychotherapies are often meant to materialize post-treatment as opposed to during (63, 64); the authors thus suggested consideration of extended counseling or "booster sessions" for maintaining treatment benefits. The investigators also found no evidence for long-term efficacy of acamprosate relative to naltrexone or placebo in terms of PDA, time to relapse to heavy drinking, and good clinical outcomes.

Gueorguieva (2010) led a recent trajectories re-analysis of the COMBINE results. The advantage of this approach over the more traditional analyses employed in the original COMBINE study (53) includes the ability to use daily drinking data and identify subgroups of subjects who show distinct patterns of clinical outcomes that might not have been hypothesized *a priori*. This trajectory approach may be more statistically powerful as well, since trajectories capture information on frequency of drinking, duration of abstinence and rate of change over time as opposed to using summary measures of drinking outcomes.

The primary aim of the COMBINE trajectories re-analyses was to estimate distinct trajectories of any/heavy drinking in addition to the effects of treatment condition on these trajectories. Gueorguieva et al. uncovered trajectories similar to those discovered in the re-analyses of the Krystal, 2001 and O'Malley, 2007 studies: abstainers (48.2% of sample), sporadic drinkers (35.5%), and consistent drinkers (16.3%); these results regard "any drinking" as opposed to "heavy drinking," although results for heavy drinking are similar. In terms of treatment effects, there were no significant effects of any treatment on the "any drinking" trajectories, although results for "heavy drinking" trajectories showed a borderline significant naltrexone by CBI interaction ($p = 0.08$).

Post hoc tests for "heavy drinking" trajectories confirmed that among those not receiving CBI, receiving naltrexone (compared to not) increased the odds of being in the "abstainers from heavy drinking" class compared to the "consistent heavy drinkers" and "sporadic heavy drinkers" classes (OR = 1.86, 95% CI (1.10, 3.16) and OR= 1.48, 95% CI (1.02, 2.14) respectively). In addition, among those not receiving naltrexone, those in CBI (compared to not) had significantly higher odds of being in the "abstainers from heavy drinking" class compared to the "consistent heavy drinkers" class (OR = 1.81, 95% CI (1.07, 3.06)). Finally, participants receiving the combination of naltrexone and CBI (referred to now on as "the Combination") compared to those receiving neither were significantly more likely to be "abstainers from heavy drinking" than "consistent heavy drinkers" (OR = 2.04, 95% CI: (1.18, 3.53)). No effects were found for acamprosate, either alone or in combination (62).

In summary, results from secondary analyses corroborated those from the original COMBINE study analyses. Primary analyses yielded no significant differences across the naltrexone, CBI, and naltrexone + CBI arms, and though naltrexone alone and placebo + CBI significantly improved outcomes, the combination was no more efficacious than either alone. Results from Donovan et al. (2008) confirmed CBI and/or naltrexone efficacy for improved outcomes beyond the active-treatment phase. Gueorguieva et al.'s *post hoc* tests further supported original COMBINE results that among those not receiving CBI, naltrexone increased odds of being in the abstainer class vs. other drinking trajectory classes. In addition, among those not receiving naltrexone, CBI increased odds of following the abstainer trajectory, and the combination of naltrexone + CBI compared to neither increased odds of following the abstainer trajectory vs. other trajectories.

Contrary to expectation, no treatment combination including acamprosate improved outcomes significantly in original COMBINE analyses (53). Donovan et al. (2008) and Gueorguieva et al. (2010) reported similar null results.

RESEARCH QUESTIONS SUGGESTED BY COMBINE RESULTS (DISSERTATION RATIONALE)

Criticisms of COMBINE and calls for causal chain analyses

The COMBINE study data are unique because naltrexone, CBI, and naltrexone + CBI were all effective, laying the groundwork for causal pathway assessment; although null effects are not exempt from mediation analyses (65), examining direct and indirect effects makes more intuitive sense for treatments that have shown to be effective. Furthermore, the collection of data across multiple time points allows for proper direct and indirect effect analysis, as the causal chain assumes the mediators succeeds treatment and precedes outcome. Finally, no one has examined the mediating variables of the COMBINE interventions, meaning that their causal mechanisms are yet to be explained.

Notably, the fact that the combined CBI + naltrexone treatment offered no advantage was unexpected (62, 66). In addition, the seeming lack of an acamprosate effect – both as a monotherapy and in combination with CBI and/or naltrexone – was particularly surprising given the positive results from European studies of acamprosate (7, 9, 16) and of the acamprosate + naltrexone combination (59).

In a 2008 series of commentaries published in the journal *Addiction*, a group of scientists offered their opinions on the COMBINE study results, as well as the COMBINE group's presentation of the overall findings. Anders Bergmark began the series by stating that the COMBINE group evaded discussion of the fact that none of their primary hypotheses were supported by the study results and provided no guidance for how future research should build on their findings. Bergmark also stated that

> …the major absence in the discussions of the results from the COMBINE study is… any discussion of the treatment mechanisms that are supposed to have generated the improvement in the participants' drinking practices: what importance should we attribute to common factors that are embedded in the treatment context and what weight should we ascribe to the pharmacologic interventions and the specific behavioral intervention? (66)

Bergmark further suggested that the absence of active acamprosate mechanisms may explain the surprising lack of acamprosate effect, and urged exploration of the "conceptual foundation of treatment mechanisms" (66).

Statements from Bühringer and Pfeiffer-Gerschel echoed those of Bergmark. They claimed that "the authors of the COMBINE study did not really state and discuss the complete failure of their theoretical study concept… This lack of concern is indeed astonishing," and that "the conclusion to recommend solely naltrexone for implementation in health care is more than questionable" (67). Bühringer and Pfeiffer-Gerschel also strongly advised studying mechanisms of change, claiming that

> The COMBINE study is outstanding for its lack of knowledge on the mechanisms of change responsible for onset, course and cessation of substance use disorders.

> Therefore future research should concentrate on mediators of change, for both behavioural and pharmacologic interventions and possible interactions. It can be assumed that not one single process, but a complex interaction of mediating and interacting contributors, is responsible for any change of behaviour. Our models of change need to be adapted accordingly. Given the present state of knowledge, the classic black box randomized controlled trial will not help us to gain urgently needed information… without an understanding of the moderators and mediators of change (67).

Amidst the arguments urging elucidation of the COMBINE results, Kalant contended that the analyses performed in the original COMBINE study did not include the no pill/CBI group because the statistical tests were *a priori* designed to test for interaction effects between medications and CBI (68). According to a 2010 reanalysis (69) that included the no pill/CBI group, all treatments were significantly better than the no pill/CBI group, but not significantly different from each other, implying that acamprosate was not necessarily ineffective in the COMBINE study (68, 69). Although Bergmark consented this may be true, he concluded the 2008 *Addiction* series stating that exploration of treatment mechanism is the necessary next step for advancing treatment (66).

Theories regarding why combination was not more effective

The lack of improvement offered by the combination of naltrexone and CBI may be explained by two phenomena. First, naltrexone and CBI may be working through the same mechanisms. For example, perhaps both naltrexone and CBI are effective because they reduce craving, and craving reduction is crucial for achieving abstinence. In this case, adding one therapy to another would not improve abstinence odds because only one therapy is necessary in reducing craving which subsequently enhances abstinence odds.

Secondly, a threshold effect may be present in that one therapy (e.g., naltrexone) reduces craving to some threshold value (e.g., none) and CBI is ineffective in patients without craving. In this second scenario, CBI's effect does not necessarily have to be mediated or explained by craving. Instead, there may be some other mechanism at work, such as reduced stress, in those without craving.

Rationale for studying craving as mediator

Although COMBINE investigators expected distinct mechanisms of action for each condition (52), craving appears to be a plausible common mediator of acamprosate, naltrexone, and CBI, as well as their combinations. Physiologic changes occurring from chronic alcohol consumption can result in craving, one behavior that characterizes addiction (1). Furthermore alcohol consumption is positively reinforced through anxiolytic and sedative effects, causing the development of alcohol-seeking behavior. Alcohol-seeking behavior, or craving, often precedes relapse, but can differ according to factors such as stress, genetic-predisposition, or social context (70).

From a mechanistic standpoint, acamprosate, naltrexone, and cognitive behavioral therapy are all alleged to affect craving in some way. Acamprosate may prevent conditioned responses to withdrawal, subsequently attenuating negative reinforcement from withdrawal symptoms by counterbalancing changes in the NMDA receptor network (70). Hence acamprosate is thought to work by reducing the craving to drink in order to abolish withdrawal symptoms. On the other hand, naltrexone suppresses the positive reinforcement from drinking: because alcohol induces the release of endogenous opioids which reinforce the pleasure of drinking, and because craving cues also activate the endogenous opioid system, naltrexone may reduce craving by blocking the conditioned activation of endogenous opioids (70). Although not biologically obvious, CBI might alleviate craving via cognitive and behavioral coping skills, enhanced self-efficacy, or increased motivation for abstinence (23). Furthermore, according to COMBINE authors craving was expected to relate directly to abstinence in addition to all treatments, including MM and CBI, as both include craving components/modules. COMBINE utilized the Penn Alcohol Craving Scale to measure alcohol craving (71).

The rationale for studying craving as a mediator is additionally supported by clinical trial results showing craving to be an important mediator of treatment outcomes. First, both landmark naltrexone studies demonstrated that naltrexone reduced craving levels (24, 25). O'Malley et al. showed that patients in the naltrexone/coping skills condition reported lower levels of craving than subjects in the placebo/coping skills therapy condition; however, placebo/supportive therapy patients also reported craving levels lower than placebo/coping skills condition patients, implying that supportive therapy may offer some advantage over coping skills in terms of craving alleviation. Patients with the highest levels of baseline craving also appeared to benefit especially from naltrexone (72). Volpicelli et al. (1992, 1995) similarly noted that patients with higher baseline levels of craving (measured by a 10-point rating scale) seemed to benefit more from naltrexone than those with lower baseline cravings.

O'Malley also found that patients treated with naltrexone reported fewer cravings before, during, and after a relapse (26). Treatment groups responded differently when asked why they stopped drinking: the naltrexone group gave reasons related mostly to lessened incentive to drink (i.e., decreased cravings or a desire to stop drinking) while the placebo group listed adverse consequences from drinking (e.g., "guilt and hangover") (26).

In a placebo-controlled, double-blind trial of naltrexone, Monterosso et al. reported that strong craving or urges to drink were associated with good clinical response to naltrexone, as indicated by the presence of a significant interaction ($p = .02$) between naltrexone treatment and baseline craving (73). Monterosso et al. measured craving with the Penn Alcohol Craving Scale (PACS), a five-item questionnaire that measures intensity, duration, and frequency of craving for alcohol during the prior week. Data from Monterosso et al. (2001) also suggested patients with high familial loadings for alcohol problems (i.e., the patient's father plus one or more additional biological relatives having alcohol problems) may also benefit especially from naltrexone.

Heilig (2006) stated that because endogenous opioids modulate the positive reinforcement of alcohol consumption, naltrexone would be expected to benefit subjects whose disease is mostly characterized by reward craving. Although this statement, along with the studies reviewed above, suggests that craving levels may *moderate* naltrexone

efficacy, as the extent of the naltrexone effect appear to be influenced by craving levels, this relationship also points to potential mediation. Identifying important moderators often elucidates potential mechanisms of action, and *variables identified as moderators may also function as mediators* (74).

Although less supported empirically, acamprosate is thought to influence craving levels theoretically. Results from Spanagel (1997) showed that acamprosate has "anti-craving properties," and that it may aid in recovering from alcohol deprivation symptoms that occur after long-term drinking. Nutt (1999) reported that because acamprosate may modulate NMDA receptor function, it could theoretically attenuate reactions to "conditioned cues" of drinking, leading patients treated with acamprosate to respond less to alcohol–related cues, such as familiar places (12, 75).

A review of the clinical aspects of craving led by Addolorato (2005) also concluded that acamprosate could target craving since the glutamatergic system, mainly the NMDA receptors, contributes to craving and subsequent relapse that may be explained by glutamatergic dysregulation related to stress from withdrawal. Hyperglutamatergic dysfunction coincides with "relief craving," or the urge to alleviate withdrawal symptoms. Because acamprosate modulates glutamate levels, relief craving is a logical mechanism of action. Mann et al. (2008) claimed that the lack of acamprosate efficacy observed in COMBINE may be due to the extremely low rates of severe withdrawal symptoms among COMBINE participants; if acamprosate does work via relief craving reduction, then the exclusion of patients with relief craving may reflect an absent biological target. Animal and human studies have also shown associations between acamprosate and changes in craving-related hormones (76). The null results found for acamprosate may reflect a flaw in the COMBINE study design.

A major distinction between CBI and the COMBINE pharmaceuticals is the proposed mechanism of action. While the literature consistently highlights craving reduction as a primary mediator of both acamprosate and naltrexone efficacy, CBI's causal chain is less understood. CBI was developed for COMBINE, precluding past examination of potential mediators. As noted above, CBI merges aspects of the Project MATCH MET, TSF, and CBT therapies. Hence mediators studied by MATCH investigators (e.g., AA involvement, noted above) are logical candidates. Potential mediators should be considered relative to CBI's specific theoretical foundations as well. For example, Cisler (2005) reported that CBI focuses on several aspects of inter- and intrapersonal functioning, such as social networks and mutual-help group attendance, dimensions that naltrexone and acamprosate do not address. Therefore potential mediators of the CBI effect include social support for sobriety and self-help group attendance, as well as personal improvements in self-efficacy or mood. CBI's pool of potential mediators thus differs dramatically from those of naltrexone and acamprosate.

However, although not biologically obvious, CBI may also affect abstinence via craving reduction. First, COMBINE authors explicitly stated that CBI addresses alcohol-related outcomes such as craving (71). Next, some consider craving to function according to cognitive and behavioral processes (23), which CBI addresses. For example, craving may result from conditioned response to prior hedonic gratification, or, at the other extreme, withdrawal symptoms (e.g., relief craving). Ultimately, cognitive and emotional states influence craving levels, as do both positive and negative cues, mental processes, mood, and self-efficacy (23). Furthermore, abstinence functions according to coping and

self-efficacy; CBI explicitly addresses these individual characteristics, implying that CBI can affect craving, either directly or indirectly via increased coping skills or self-efficacy. Finally, according to COMBINE authors, next steps for analyses require assessing combinations of pharmacotherapies and CBI and their effects on craving reduction (71). *The importance of craving as a potential mediator of CBI efficacy lies in its candidacy as the only logical common mediator across the three effective COMBINE therapy combinations.*

In the 2008 *Addiction* commentary series, Bergmark asked, "Is it possible that the lack of combination effects is related to the fact that the stipulated neurochemical mechanisms are not really empirically identified, or not connected significantly to the issue of alcohol craving?" (66). Analyzing the causal mechanisms of the COMBINE conditions will provide an answer to Bergmark's question.

Rationale for studying stress as mediator

Although COMBINE investigators did not hypothesize that stress would mediate treatment effects, a primary goal of behavioral interventions is to teach skills for coping with stress without resorting to drinking. Because COMBINE does not have a measure of "coping skills," stress was examined instead. This decision was informed by Susan Folkman (personal communication).

SUBSTANTIVE BACKGROUND SUMMARY

Acamprosate and naltrexone have exhibited clinical efficacy in numerous trials. Although Project MATCH investigators could not empirically establish that their MET, TSF, and CBT psychotherapies were superior to a control condition, all treatments did appear to improve drinking outcomes, with a slight overall advantage observed for TSF. COMBINE investigators chose acamprosate and naltrexone based on their evident efficacy and developed CBI according to the strongest elements of the Project MATCH psychotherapies.

COMBINE investigators sought to establish whether combining pharmaceuticals and behavioral treatments for alcoholism could improve outcomes beyond any individual treatment. Contrary to expectation, COMBINE investigators found that although naltrexone improved outcomes without CBI, and CBI improved outcomes without naltrexone, *the combination of naltrexone + CBI was no more efficacious than each individually*. Yet studies of naltrexone and behavioral therapy suggest that naltrexone and behavioral therapy work through different causal mechanisms. If this were true, then why is the effect of the COMBINE naltrexone + CBI arm not stronger than either naltrexone or CBI administered individually? For my dissertation, I will describe the causal pathways of the COMBINE study treatments and identify mediating variables, as well as investigate a possible treatment threshold effect in order to help explain the lack of improvement offered by the combined naltrexone and CBI treatment. The overall goal of my dissertation research is to identify the causal mechanisms of the COMBINE naltrexone, CBI, and naltrexone + CBI treatments, as well as to estimate the extent to which these treatments affect drinking outcomes through various intermediate variables. Because I

hypothesize each treatment affects several causal paths, I propose to examine controlled and natural direct effects as well as natural indirect effects in the context of various potentially mediating variables. Because craving may be the only common mediator across acamprosate, naltrexone, and CBI conditions, my primary aims revolve around examining causal pathways involving craving.

II. METHODS BACKGROUND: APPROACHES TO QUANTIFYING MEDIATION

The classic Baron and Kenny method

Baron and Kenny's seminal mediation/moderation paper (1986) outlines both mediation and moderation models, and describes the now eminent method of identifying the presence of a mediating, or intervening, variable (74).

According to Baron and Kenny, for mediation to be present the following conditions must hold:

1. Variations in levels of the IV account for variations in the mediator.
2. Variations in levels of the mediator account for variations in the DV.
3. When paths from the IV to the mediator and the mediator to the DV are controlled, a previously statistically significant relationship between the IV and DV is no longer significant.

Baron and Kenny also present steps to assess the above conditions and determine whether mediation exists:

1. Regress the mediator on the IV to show the IV significantly affects the mediator.
2. Regress the DV on the IV to show the IV significantly affects the DV.
3. Regress the DV on the mediator and the IV to show the mediator significantly affects the DV when controlling for the IV.

4. Show the effect of the IV on the DV is less in the third regression than in the second.

Perfect mediation exists when the IV has no effect on the DV once the mediator is included in the regression equation. Note that the third regression will have less power due to the assumed multicollinearity of the IV and the mediator (since the IV is assumed to predict the mediator).

Although the Baron and Kenny method generalizes for use with binary and categorical outcomes and mediators, the differences between linear and logistic regressions have important implications for implementing this strategy. The scale of the dependent variable in linear regressions is constant across the regression equations, while the scale in logistic regressions varies according to the percent of explained variance, which depends on the covariates included in each equation. Therefore when analyzing binary outcomes, the scale of the outcome Y varies according to the explanatory variables; this implies that the scale of Y in Y~X, M differs from those in both Y~X and Y~M. In other words, the coefficients produced in the three regressions cannot be compared meaningfully without standardization. Standardizing the regression coefficients is imperative to proper analysis of binary outcomes and mediators. Standardization simply requires multiplying each coefficient by the standard deviation of the IV and dividing by the standard deviation of the DV.

To calculate the extent of mediation, we would simply take the ratio of the indirect effect (the product of the path coefficient from exposure to mediator * the path coefficient

from mediator to outcome) to the total effect (the path coefficient from exposure to outcome when not controlling for mediators) (74).

Alternatives to Baron and Kenny

Due to its simplicity and elegance, the Baron and Kenny method is the most widely known and applied method for establishing mediation in the social sciences (77). However, it is not necessarily the best. MacKinnon, et al. describe several lesser-known methods that may perform better under certain conditions. The authors compare the statistical performances of 14 methods to assess mediation. MacKinnon notes that these 14 methods are quite diverse and therefore differ in their null hypotheses, assumptions, and statistical methods (77).

MacKinnon presents the methods in three categories. The first category, "causal steps," includes the Baron and Kenny method as well as the Judd and Kenny and the "joint significance" methods; the authors label these "causal step methods" due to their foundations in serial tests of the links in causal chains. The second group of tests (Freedman & Schatzkin, McGuigan & Langholtz, Clogg, et al., and Olkin & Finn) is based on the differences in the regression coefficients before and after adjusting for the mediating variable; these tests are hence called the "difference in coefficients" tests. The third and final group of tests (Sobel, Aroian, Goodman, MacKinnon, et al. ("distribution of products"), MacKinnon, et al. ("distribution of $\alpha\beta/\sigma_{\alpha\beta}$"), MacKinnon & Lockwood, Bobko &Rieck) is based on examining the product of the coefficients from the paths in path models.

MacKinnon claims that the first group, which includes the famous Baron and Kenny method, usually has the most limitations. He claims that the Baron and Kenny and Judd and Kenny methods establish conditions for mediation, but do not statistically test for indirect effects (i.e., the mediated effects of the IV on the DV) because they do not include joint tests of the three relationships, and because there is no way to tell whether the direct and indirect effects cancel each other out (i.e., a rare but possible situation that the mediator suppresses the effect of the IV).

MacKinnon suggests using tests from groups two and three. The second group of methods involves testing whether corr(IV, DV) differs from the partial corr(IV, DV | mediator) by testing the significance of the difference in coefficients before and after adjusting for the mediator. The third group, "product of coefficients," tests the significance of the mediator by dividing the product of the effect estimates of the mediator and exposure (i.e., multiply each of the effects on the outcome) by their respective standard errors and comparing it to the standard normal distribution (i.e., using z-scores).

MacKinnon then calculates the power (proportion of times the method leads to rejection of false null hypotheses) and type I error rates (α , proportion of times method leads to rejection of true null hypothesis) of the 14 tests (three groups) via simulations. He finds that most tests have minimal bias except for the first MacKinnon, et al. test. He concludes that the causal steps tests have low power for small and medium effects, and α's below the nominal value for all sample sizes; the joint significance test has the highest power and lowest α of the causal steps methods; most of the difference in coefficients methods have high power ($\geq .80$), but only the Clogg, et al. and Freedman & Schatzkin methods have accurate α's; the product of coefficient methods have the most accurate α's

and most power of all the tests. He suggests that the first and second MacKinnon methods, the Clogg method, and the Freedman & Schatzkin method perform the best, but warns that Clogg assumes fixed effects, which conceptually may not make sense.

MacKinnon closes with statistical recommendations for testing mediation. He says that studies using causal step methods are the most likely to miss real effects (because they have the lowest power) but are very unlikely to commit type I errors. Furthermore, the difference in coefficients methods are the most powerful and have the most accurate α's (under the null hypothesis), yet often inaccurately estimate standard errors. The product of coefficients methods are powerful as well, but have very low α's; furthermore the distribution of coefficients method is the most accurate of these methods (under the null) and has the most power.

MacKinnon notes that the choice of method depends on the null hypothesis, which can address either β coefficients or correlation coefficients, as well as considerations of power and type I error. For example, if we were interested in the percent of the exposure's effect that is mediated, the best choice might be a product of coefficients method to calculate the path coefficients. We would then take the difference between the mediator~exposure coefficient and the outcome~mediator coefficient, and divide that quantity by the outcome~exposure coefficient (condition on the mediator) to arrive at the size of the indirect effect. On the other hand, if we were interested in correlation coefficients, we would use a different approach, such as partialling correlations.

Additional issues in establishing mediation

Multiple mediation

Preacher and Hayes argue that analyzing several mediators simultaneously offers many advantages over tests which examine potential mediators individually (78). First, the multiple mediator test is similar to performing a regression with several covariates, the goal being to establish an effect; if the effect is significant, then we would attribute it to the set of predictors (mediators, in this case) as a whole. This is intuitively appealing since we would expect that an exposure would impact more than one intervening variable, as treatments are often designed to do. The second advantage of multiple mediation tests is the ability to see the extent to which each potential mediator mediates the exposure's effect on the outcome. Third, the estimate of the effect is less likely to be biased due to omitted variables because we are able to control for several mediators simultaneously; when examining potential mediators individually, the parameter estimate is more prone to bias because of the omitted variable problem. Controlling for multiple mediators at once lowers the probability of mistakenly attributing effects to one mediator when in fact the effects are partially explained by another variable/mediator. Finally, examining multiple mediators simultaneously allows us to compare the magnitudes of the indirect effects specific to each mediator via contrasts. Therefore the investigator can see how the total indirect effect of the exposure on the outcome decomposes among the proposed mediators, as well as judge the specific effects of each mediator while controlling for all others (78). Although traditional approaches such as the Baron and Kenny method cannot handle assessment of multiple mediators simultaneously, the causal inference methodology (described below) is equipped for addressing questions regarding "multiple mediation."

Bootstrapping

Bootstrapping the standard error/confidence interval of the? indirect effect is appealing for various reasons. First, the distribution of the products of coefficients (e.g., Sobel's statistic) tends to be positively skewed; thus imposing a normal distribution and calculating traditional confidence intervals is usually inappropriate. Furthermore, we know that the product of two normal variables is not normal (77). Violating the normality assumption can lead to underpowered tests. MacKinnon (2002) and Preacher and Hayes (2004), as well as van der Laan (personal communication), therefore suggest bootstrapping the indirect effect to attain proper confidence intervals and avoid problems with power.

Bootstrapping requires the creation of *B* samples of size *n* from the original sample, with replacement. This means that one case can reappear in the new "resamples" several times, or not at all. If we were to bootstrap a product of coefficients estimate, for example, we would estimate the products of the path coefficients from the *B* resamples. This involves repeating the estimation process *B* times (where *B* is usually 1000 or more), and obtaining *B* estimates of the direct and indirect effects of the treatment on the outcome. The resulting distribution is a non-parametric estimate of the distribution of the mediated effects. The 100(1-α)% bootstrap confidence interval is then computed by ranking the *B* estimates from low to high and assigning the lower and upper 100(α/2)% values of the ranked list as the lower and upper 100(1-α)% confidence bounds, respectively. For example, if $B = 1{,}000$ and $\alpha = 0.01$, we would take the 10^{th} and 991^{st} values from the ranked list as our respective lower and upper 99% confidence limits.

Bootstrapping is far superior to traditional confidence interval estimating methods that predicate normality of the sampling distribution of the indirect effects (77). Because there are several bootstrap confidence interval options, such as the percentile, bias-corrected (BC), and bias-corrected accelerated (BCa) intervals, the researcher must choose which to implement. Although the computational differences among the three bootstrap confidence intervals will not be discussed here, Preacher and Hayes (2004) and MacKinnon (2002) recommend either the BC or BCa because of power and Type I error issues. MacKinnon, et al. also suggest using the BC confidence interval because it had the most power and best performance in terms of robustness in a simulation study performed by the authors.

Limitations to mediation analyses

As with any analysis, measurement error will bias the mediated effect estimates (77). Any reciprocal effects among mediating variables can also lead to bias. Another problem stems from the expected collinearity of multiple mediating variables; since the mediators are assumed to have a common cause (e.g., treatment), they are inherently collinear, which inflates the coefficients standard error and in turn lowers power. Related to collinearity is the extremely problematic assumption that the treatment and mediator do not interact; most traditional approaches to assessing mediation make this assumption even when empirical evidence shows it to be untrue. Finally, we must assume the causal chain from the exposure to the mediator to the outcome is in fact real, and that we have

not omitted any variables. Excluding variables that have causal links to either the outcome or any of the mediators may also bias estimates.

Examples of the use (and misuse) of mediation methods in alcohol research

In the alcohol field, several studies have focused on mediators of treatment and other exposures that aim to increase abstinence from alcohol and drugs. A cursory review of this body of literature (as directed by Dr. Kaskutas) shows the occasional misuse of mediation methods, as well as the frequent misinterpretation of results. Numerous authors have implemented the Baron and Kenny strategy without standardizing the logistic regression coefficients or addressing the method's lack of a joint significance test. Among the investigators who utilize Sobel's test instead, few bootstrap the indirect effect or even mention the normality assumption. Perhaps due to computational difficulties, multiple mediation analyses are rarely undertaken; those investigators who have attempted multiple mediation analyses misinterpreted the results.

Kaskutas, et al. test whether the relationship between Alcoholics Anonymous involvement and alcohol problem severity is mediated by changes in social network composition (79). Following the Baron and Kenny method, Kaskutas compares the beta coefficients of the path models with and without the mediating paths by using t-tests to determine if the betas differed significantly, and to gauge the relative size of the mediating effect. She performs linear regressions to show that AA involvement predicts problem severity, and that number of drinking influences along with social network size predict problem severity. Kaskutas concludes that the effects of AA involvement on drinking outcomes decrease by 36% once the mediating path is included in the path model.

The main problem with Kaskutas' conclusions is that she fails to establish a temporally sound causal pathway from exposure→mediator→ outcome; Kaskutas uses social network composition and problem severity measured at the same time point, but claims that a large social network *causes* better drinking outcomes. Although the importance of establishing temporality can be difficult, Kazdin and Nock argue it is the foundation of mediation analyses (80). Others say having a cause temporally precede an effect is simply more intuitively appealing.

In a similar study of adolescents in AA, Kelly, et al. hypothesize that self-efficacy, motivation, and coping at three months after treatment will mediate the effects of 12-step affiliation on substance use in the subsequent three months (81). Kelly also implements the Baron and Kenny method, fitting structural equation models (SEM's) to determine path coefficients, and because of the rather small sample (n=74), runs ordinary least squares regressions to replicate the results.

Kelly finds that the relationship between each of the three mediators and 12-step attendance becomes insignificant once affiliation is included in the model; Kelly claims that 12-step affiliation therefore mediates the effects of 12-step attendance on motivation, coping, and self-efficacy. However, all mediators and outcomes are measured at three months, so inference of mediation from this particular analysis is, again, debatable. Furthermore, because of evident suppression of motivation on affiliation, Kelly "adjusts downward." What this means is unclear, yet he suggests that the path model post-downward adjustment shows motivation partially mediates the effects of affiliation on use. Kelly then says that "formal mediation analyses" demonstrate complete mediation by

affiliation on attendance's effects on motivation because affiliation explains unique variance of motivation; however, he does not explain what "formal mediation analyses" are. Kelly's conclusions may be inaccurate because of his ambiguous methodology. Although he proposes that coping, motivation, and self-efficacy mediate the effects of affiliation on substance use, his conclusions focus on the mediation effects of affiliation.

In a study of military personnel, Williams, et al. implement a Preacher and Hayes multiple mediation approach (78) to determine whether particular intervening variables mediate the effects of participation in a web-based intervention designed to curb problem drinking (82). The sample consisted of 3,889 participants randomized to undergo Alcohol Savvy (AS), Drinker's Check-Up (DCU), or neither. Both AS and DC, which are interactive, web-based programs, were shown to reduce heavy drinking at the one-month follow-up. The authors offer four general categories of potential mediators of this effect: perceived social norms, concern about drinking, readiness to change, and stress management.

Because mediators and outcomes are measured at both the one-month and six-month follow-ups, Williams, et al. perform two sets of mediation analyses, one with lagged mediators and one with concurrently-measured mediators. Williams, et al. use the Preacher and Hayes multiple mediation macro to determine whether the four groups of potential mediators do in fact intervene upon the effects of AS, DCU, or both. They find that in the concurrent model, both AS and DCU affect perceived norms, and that perceived norms in turn affect alcohol use. Although neither DCU nor AS appear to affect any of the other potential mediators mentioned, all mediators significantly predicted drinking outcomes. The results of the lagged model are similar.

Although the study of Alcohol Savvy and Drinker's Check-Up is a prime opportunity to utilize the multiple mediation approach, the analyses appear problematic. The authors interpret the significance of perceived norms in the multiple mediation model as simply a "mediated effect" and fail to account for the fact that the model includes several mediators simultaneously. The effect of perceived norms is actually a specific indirect effect, which reflects its ability to mediate the effects of the intervention *conditional on the inclusion of the other mediators*. This condition, which Williams, et al. did not appear to consider, is critical to correct interpretation of the results of the multiple mediation analyses.

Moderation, mediation and indirect effects in the epidemiologic literature

Although the issues regarding mediation in epidemiology are conceptually similar to those in the social sciences, the terminology is somewhat different. For example, epidemiologists seem to prefer the term "intermediate variable" to the social scientists' "mediator." Here I use the terms interchangeably. Furthermore, the methods suggested in the epidemiologic literature are distinct from those previously discussed. VanderWeele emphasizes that the traditional social science methods can only be used in the absence of a mediator*outcome interaction, or moderation of treatment effects by the mediator; this point is rarely addressed in the social science mediation literature (83).

Baron and Kenny define moderators as third variables that establish the maximum effectiveness of the treatment relative to the dependent variable. Moderators affect the direction and strength of the independent and dependent variables' relationship, implying

a change in the correlations between the independent variable (IV) and dependent variable (DV) according to the level of the moderator.

The proper analyses for determining moderation depend on whether the IV, DV, and moderator are dichotomous or continuous variables. Baron and Kenny illustrate the various moderation analyses for all combinations. For example, if all three were dichotomous, we would demonstrate moderation via ANOVA's; moderation would manifest as an interaction between the IV and the proposed moderator. If one or more of the three variables were continuous, the analyses become more complicated, requiring regressions, testing differences in correlations, and potentially dichotomizing the continuous variables on logical cut points (e.g., points of acceleration in a step function). Moderators are present when there is an unexpected or inconsistent relationship between the IV and DV (e.g., when a relationship holds in one subpopulation but not another), and ultimately account for some extent of the relationship between the IV and DV. On the other hand, a mediator is the mechanism through which the treatment acts upon the outcome. Moderators tell us *when* certain effects exist, whereas mediators tell us *why*.

When a moderator is also a mediator, the presence of a mediator*outcome interaction biases the results of structural equation modeling and path analyses, which are the methods used in the aforementioned alcohol research examples. The methods suggested by VanderWeele (2009) and Petersen (2006) can estimate direct and indirect effects when a mediator*outcome interaction exists, a primary advantage of the methods discussed below. However, if the goal to estimate the causal effect of an exposure while holding the level of the mediator at a controlled level defined by the researcher (controlled direct effect, discussed further below), multivariable regression can yield unbiased estimates, assuming no unmeasured confounding and the absence of any confounding of the mediator that is itself affected by the exposure (83, 84).

Controlled and natural direct effects

A controlled direct effect is the effect an intervention has on an outcome while the mediator is fixed to some pre-determined level; hence we assess the direct effect of the intervention while *controlling* the level of the mediator according to a pre-specified level. In counterfactual terms, the controlled direct effect is the difference in the counterfactual outcome Y_a if the individual were unexposed and his mediator value was set to $M=m$ compared to his counterfactual outcome if he were exposed and his mediator value remained at the same level $M=m$. When the exposure and the mediator do not interact, controlled and natural direct effects are the same. Petersen, et al. argue that controlled direct effects can be estimated with traditional multivariable regressions, even when the exposure and mediator do interact; in this scenario, the number of controlled direct effects equals the number of levels of the mediator variable (84).

Similarly, a natural direct effect is the effect the intervention has on the outcome when the level of the mediator is set to whatever it would have been in the absence of the intervention; here we are interested in the effect of the intervention while allowing the level of the mediator to be what it *naturally* would have been without the intervention. In counterfactual language, the natural direct effect is the difference in the counterfactual outcomes when an individual is exposed versus unexposed, while the level of the mediator retains the value it would have without exposure. In the absence of an exposure*mediator

interaction, we can again use standard multivariable regressions to estimate the natural direct effect which is, in this case, equivalent to the controlled direct effect. Furthermore, as opposed to estimation of the controlled direct effect which results in as many estimates as there are levels of the mediator, the natural direct effect calculation results in a single estimate of the population-level direct effect.

Several assumptions must be met to accurately determine the presence of either controlled or natural direct effects. Estimation of controlled direct effects requires:

1. No unmeasured confounding of the effect of the exposure on the outcome.
2. No unmeasured confounding of the effect of the mediator on the outcome.

Estimation of natural direct effects requires these 2 assumptions as well as:

3. No unmeasured confounding of the effect of the exposure on the mediator.
4. Within subgroups defined by covariates, the value of the mediator without exposure gives no information about the expected strength of the exposure's effect on the outcome when the mediator is controlled. In other words, within subgroups defined by baseline covariates, counterfactual outcomes are independent of the value of the mediator in the absence of exposure.

Finally, the use of standard multivariable regressions to assess either controlled or natural direct effects further requires:

5. No confounding or interaction by causal intermediates; i.e., the exposure cannot affect covariates which confound the effect of the mediator on the outcome.

In situations where confounders are affected by the exposure, traditional regressions will yield biased results, since the exposure's effect will partially be controlled via control of the confounder.

Petersen outlines direct effect analysis according to the following steps:

1. Regress outcome on confounders, exposure, and mediator.
2. Estimate E(Y|A=1, M=m, W) - E(Y|A=0, M=m, W) using the model from step 1. This yields the controlled direct effect.

3. Regress outcome on confounder while fixing each individual's mediator to an estimate of the value it would be if the individual had been unexposed. This yields the individual natural direct effects.

4. Regress mediator on exposure and confounders to estimate the expected value of the mediator in the absence of treatment ($E[Z_0]$).

5. Repeat step 4, but include weighting by confounders (W). This yields an estimate of ($E[Z_0*W]$).

6. Substitute $E[Z_0]$ (from step 4) and $E[Z_0*W]$ (from step 5) into the model produced in step 3. This yields the population-level natural direct effect.

This effect can be interpreted as the effects of exposure vs. non-exposure, if the effects of the exposure on the mediator were blocked (84).

Causal inference methods and direct effect estimation

A key difference between traditional mediation approaches and the causal inference approaches outlined below is this: instead of taking a common tool (regression) and using it to try to assess mediation, causal inference approaches start with clear definition of the parameter of interest using the counterfactual framework. Then various approaches can be used to estimate that parameter. I chose to use the causal inference approach because the emergence of causal inference approaches will help researchers circumvent the problems posed by traditional mediation methods, such as the Baron and Kenny method, by both providing a framework for making assumptions and definition of the causal parameter explicit while also promoting data-adaptive, non-parametric approaches to optimize efficiency and control of confounding.

Using causal models requires several steps, with the ultimate goal being to use the observed data to map an algorithm to a parameter that provides a summary of the data-generating distribution. Causal parameters summarize how the distribution would change if experimental conditions (e.g., treatment, mediators) changed. However, this inference requires a set of assumptions regarding the data-generating distribution (discussed below), of which only some can be tested statistically.

III. SPECIFIC AIMS

I will estimate and compare the direct and indirect effects of the COMBINE Study naltrexone, cognitive behavioral intervention (CBI), and naltrexone + CBI arms, compared to the placebo-only control group, using Targeted Maximum Likelihood Estimation (85). Specifically, my aims are:

1) To estimate treatment effects on abstinence (16-week percent days abstinent) using Targeted Maximum Likelihood Estimation.

2) To assess moderation of treatment effects on abstinence (16-week percent days abstinent) by cravings and/or stress levels (measured at 4- and 12-week follow-ups) (i.e., estimate controlled direct effects).

3) To estimate the treatment effects on abstinence (16-week percent days abstinent) attributable to pathways that do not involve cravings and/or stress (measured at 4- and 12-weeks) as their causal mechanism (i.e., estimate natural direct effects).

4) To assess the extent of the treatment effects on abstinence (16-week percent days abstinent) attributable to cravings and/or stress (measured at 4- and 12-weeks) (i.e., estimate indirect effects and calculate "proportion explained").

IV. METHODS

SAMPLE

COMBINE investigators randomized 1,383 newly abstinent participants to nine groups. Participants were predominantly white, male, middle-aged, married/in a stable relationship, relatively well educated, and employed (53, 63). Enrollment criteria included: minimum average weekly alcohol consumption of 14 drinks for women or 21 drinks for men within a 30-day period in the 90 days prior to baseline; at least two heavy drinking days in the 90 days prior top baseline; and zero breath alcohol level before signing consent. Exclusion criteria included: meeting DSM-IV criteria for any psychiatric disorder requiring medication; opiate dependence/abuse in six months prior to baseline; current drug dependence (besides marijuana or nicotine); more than seven days inpatient treatment for substance use disorders in 30 days before randomization; planned continuation in pre-occurring alcoholism treatment during treatment phase of study; or abstinence from alcohol for more than 21 consecutive days prior to randomization (53, 63).

Over a 16-week period, patients received either active/placebo naltrexone, or active/placebo acamprosate (total of 4 medication conditions: placebo, naltrexone, acamprosate, and naltrexone + acamprosate) either with or without the combined behavioral intervention (4 medication regimens x 2 CBI regimens = 8 groups). The ninth and final group underwent CBI only, without placebo or active medication.

Medical Management

	Placebo	Acamprosate
Placebo	1	2
Naltrexone	3	4

Medical Management + Psychotherapy

	Placebo	Acamprosate	No pills
Placebo	5	6	–
Naltrexone	7	8	–
No pills	–	–	9

FIGURE 1. COMBINE treatment combinations (Cell 9 patients received only CBI, no MM)

*Figure taken from Mattson and Litten, Combining Treatments for Alcoholism: Why and How? J. Stud. Alcohol, Supplement No. 15: 8-16, 2005

The active naltrexone groups received 100 mg doses, the active acamprosate groups received 3 mg doses, and those in CBI groups received up to 20 therapy sessions. The CBI only (no pills) group had access to 4 sessions with a health care professional who

provided health care advice, while all participants treated with pills (either active medication or placebo) received Medical Management (MM), a low-intensity protocol intended for use in routine clinical practice. CBI, on the other hand, reflects treatment delivered in addiction specialty programs. During the medical management sessions, a health care professional provided information about the medications and adherence, and recommended support group attendance as well as abstinence from alcohol and drug use (86). One of COMBINE's goals was to determine whether the moderate-intensity CBI improved patient drinking beyond MM (71, 87). Investigators further hypothesized that CBI would enhance pharmacologic effectiveness.

Treatment adherence had a median of 96.4% across all pill-taking groups (including placebo) and the attrition rate was 6%. Participants were interviewed 9 times during the 16-week treatment period and at 26, 52, and 68 weeks after the treatment period ended. Interviews assessed physical, physiologic, and laboratory measures; expectancies of treatment; behavioral measures; alcohol and drug involvement; craving; motivation; and psychological and psychiatric measures (53). The two primary outcomes were time to first heavy drinking day and percent days abstinent in the 16-week treatment period.

CAUSAL ROAD MAP

Van der Laan and Rose (2011) recommend outlining the analyses plan according to the following roadmap:

1. Specify the Questions
2. Specify the Causal Model
3. Specify the Causal Parameter of Interest
4. Assess Identifiability
5. Commit to a Statistical Model and Target Parameter of the Observed Data Distribution
6. Estimate the Chosen Parameter of the Observed Data Distribution
7. Interpret Results

1. Specify the Questions

Based on the specific aims, the questions here are:

1) To estimate treatment effects on abstinence (16-week percent days abstinent) using Targeted Maximum Likelihood Estimation.
2) To assess moderation of treatment effects on abstinence (16-week percent days abstinent) by cravings and/or stress levels (measured at 4- and 12-week follow-ups) (i.e., estimate controlled direct effects).
3) To estimate the treatment effects on abstinence (16-week percent days abstinent) attributable to pathways that do not involve cravings and/or stress (measured at 4- and 12-weeks) as their causal mechanism (i.e., estimate natural direct effects).
4) To assess the extent of the treatment effects on abstinence (16-week percent days abstinent) attributable to cravings and/or stress (measured at 4- and 12-weeks) (i.e., estimate indirect effects and calculate "proportion explained").
5) To assess the possibility of a threshold effect by assessing treatment effects on cravings and/or stress (measured at 4- and 12-weeks) and interpreting the results in

the context of those from the controlled direct effect analyses; e.g., perhaps one treatment affects mechanisms earlier and the other does not work once those mechanisms are affected by the first.

2. Specify the Causal Model

The non-parametric structural equation model (NSPEM), or "causal model" takes the following structure:

- A = COMBINE tx condition = (placebo, naltrexone, CBI, naltrexone +CBI)
- W = Covariates (baseline PDA, tx site)
- Z = Cravings, Stress
- Y = Percent Days Abstinent

→ $W = f_w(U_w)$
→ $A = f_A(W, U_A)$
→ $Z = f_Z(A, W, U_Z)$
→ $Y = f_Y(A, W, Z, U_Y)$
→ $U = f_U(U_W, U_A, U_Z, U_Y) \sim P_U$

3. Specify the Causal Parameter of Interest

Relating back to the specific aims:

1) To estimate treatment effects on abstinence (16-week percent days abstinent) using Targeted Maximum Likelihood Estimation.
   → Additive Treatment Effect: $E(Y_1 - Y_0)$
2) To assess moderation of treatment effects on abstinence (16-week percent days abstinent) by cravings and/or stress levels (measured at 4- and 12-week follow-ups) (i.e., estimate controlled direct effects).
   → Controlled Direct Effect: $E(Y_{1z} - Y_{0z})$
3) To estimate the treatment effects on abstinence (16-week percent days abstinent) attributable to pathways that do not involve cravings and/or stress (measured at 4- and 12-weeks) as their causal mechanism (i.e., estimate natural direct effects).
   → Natural Direct Effect: $E(Y_{1Zo} - Y_{0Zo})$
4) To assess the extent of the treatment effects on abstinence (16-week percent days abstinent) attributable to cravings and/or stress (measured at 4- and 12-weeks) (i.e., estimate natural indirect effects).
   → Indirect Effect: Total Effect–NDE = $E(Y_1 - Y_0) - E(Y_{1Zo} - Y_{0Zo})$
   $= E(Y_{1Z1} - Y_{0Zo})$

4. Assess Identifiability

The following assumptions need to be met in order to identify CDE's (83, 84).:

1) No unmeasured confounding of the effect of the exposure on the outcome.
   $A \perp\!\!\!\perp Y_{az} \mid W$
2) No unmeasured confounding of the effect of the mediator on the outcome.
   $Z \perp\!\!\!\perp Y_{az} \mid W$

To identify NDE's, two additional assumptions must hold as well:

3) No unmeasured confounding of the effect of the exposure on the mediator.

$$A \perp Z_a \mid W$$

4) Within subgroups defined by covariates, the value of the mediator without exposure gives no information about the expected strength of the exposure's effect on the outcome when the mediator is controlled. In other words, within subgroups defined by baseline covariates, counterfactual outcomes are independent of the value of the mediator in the absence of exposure.

$$((A,Z) \perp Y(a,z) \mid W \text{ for all possible } a,z \text{ values of } (A,Z))$$

$$E(Y_{az} - Y_{0z} \mid Z_0 = z, W) = E(Y_{az} - Y_{0z} \mid W)$$

5. Commit to a Statistical Model and Target Parameter of the Observed Data Distribution

I will use Targeted Maximum Likelihood Estimation to estimate the target parameter Ψ^*. The formal proof for identifiability shows that direct effects are functions of the observed data distribution , i.e., $\Psi^*_{NDE} = E_0 \Sigma E(Y(1, z) - Y(0, z)|W) g^*(z|W,0)$, where the conditional distribution of $Z|W$ is $g^*(z|W,A=0) = \{[g_0(z|W,0) - g(z|W,0)] * g_0(0|W)/g(0|W) + g(z|W,0)\}$. Because $g(A|W)$ will be correctly specified when treatment is randomized, then $\Psi^*_{NDE} = \Psi_{NDE}$, yielding a consistent estimate of the natural direct effect. This illustrates the double-robustness of the efficient influence curve since in the case of randomized control trial data, we know $g_0(A|W)$, so our estimate will be consistent if either $Q=Q_0$ or $g_Z=g_{0Z}$.

6. Estimate the Chosen Parameter of the Observed Data Distribution

Please see the "Statistical Analyses" section below.

7. Interpret Results

Please see the "Results" section below.

Notation and assumptions: putting the "Roadmap" into words

Let the full data be denoted by X. We can partition X into the intervention (treatment) variable (A) and non-intervention variables (covariates) (W). The post-treatment distribution of the outcome $Y_a(U)$ is a parameter of the distribution of (U, X). The full data $X^F = (W_a : a \in A) \sim F_X$, where A represents the set of all possible treatment levels and F_X represent structural equations coinciding with the data distribution. Defining a structural causal model (SCM) imparts the allowed distributions for the full data X (85).

The observed data, O, is assumed to be a subset of the full data X, and is defined as a multidimensional random variable $O = (W, A, Y) \sim P_0$, where P_0 is the distribution of the observed data. Because X is defined by the distribution of U and the structural equations F_X, and because O is assumed to be a subset of X, U and F_X thus define O. F_X is the distribution of X, The distribution of (U,X), as defined by the non-parametric structural equation model, implies a distribution of O since O is a function of (U,X) (even subset of

X). We assume O has the probability distribution P_0, where P_0 factorizes to $P_0(O)=P_0(W,A,Y)=P_0(Y|A,W)\ P_0(W)P_0(A|W)$.

The parameter Ψ maps the output from the statistical model to the parameter space, using the distribution assumed (if any) in the statistical model M. The statistical model M defines the set of possible probability distributions of O. P_0 is in M.

The target parameter is denoted as $\Psi_F(P_{U,X})$. We use the observed data O to estimate this parameter with $\Psi(P_0)$. Using $\Psi(P_0)$ to estimate $\Psi_F(P_{U,X})$ requires identifiability, or the assumptions needed to write $\Psi(P_0)$ as $\Psi_F(P_{U,X})$. We first must assume consistency, or that the observed outcome for a given participant is equal to its counterfactual value with treatment A set to its observed value. Formally, we must assume that $Y(U)=f_Y(A, W, U_Y) =Y_A(U)$, i.e., $Y_a \perp A|W$. This "assumption" is sometimes referred to as a theorem following logically from the counterfactual definition.

The two assumptions that we are required to make in order to identify the target parameter are coarsening at random and experimental treatment assignment. The coarsening at random assumption (CAR) states that conditional on measured confounders, treatment assignment is independent of outcomes; i.e., given the full data $X = (W,Y_0, Y_1)$, $A \perp X|W$. The experimental treatment assumption (ETA) states that within strata defined by measured confounders, the conditional probability of receiving treatment is bounded away from 0 and 1; i.e., $P(A = a|W) > 0$ for all possible a (85).

SUPER LEARNER

Estimating the target parameter, which we now denote by $\Psi(Q_0)$, involves several steps including implementation of Super Learner and Targeted Maximum Likelihood Estimation. First, we estimate Q_0 using the machine learning algorithm Super Learner. The Super Learner approached rests on three fundamental concepts: semi-parametric estimation, loss-based estimation, and cross-validation (85).

Semi-parametric estimation allows the data to inform estimates without analyst-imposed parameterization. Super Learner enables researchers to run and compare a variety of analyst-supplied algorithms, and returns an optimal prediction function with the lowest cross-validated risk; this advantage precludes both the need to commit to a single algorithm as well as biased parameterization.

Loss-based estimation means that loss functions are used to define the best estimator of the causal parameter; the "best" estimator is what which minimizes the loss function, The loss function L(Q,O), defines risk as the expectation of loss at Q, and the true Q_0 as the minimizer of risk. The loss function assigns a performance measure to a candidate function Q when Q is applied to O. For example, if we specified a squared error loss function $L(O, Q)=(Y-Q(A, W))^2$, we would use the observed data $O\sim P_0$ and choose the algorithm that minimizes this quantity. Thus we can write the parameter of interest as $Q_{0bar} = argmin_{Qbar}E_0L(O, Q_{bar})$. The optimal Q_{0bar} minimizes $E_0L(O, Q_{bar})$. The expected loss is also called the risk. We have now introduced both Q_{bar} as well as Q. $\Psi(Q_0)$ depends on $Q(W)$ and Q_{bar} and we use Super Learner to estimate Q_{bar} (85).

Choosing the algorithm that minimizes the risk demands cross-validation, which allows accurate estimation of the expected loss. Cross-validation means that the observed data are partitioned into "training sets" and corresponding "test sets"; the use of cross-

validation makes Super Learner a data-adaptive approach and permits comparison of algorithms across independent datasets from the same distribution. Cross-validation is used to select the "best" algorithm from a library of many algorithms, and ultimately to assess the overall performance of the Super Learner algorithm itself. Using cross-validation also allows us to try many algorithms without overfitting. Super Learner particularly implements *V*-fold cross-validation to ensure computational ease as well as fulfillment of finite sample and asymptotic properties. In "*V*-fold cross-validation," the observed data $O = O_1, \ldots, O_n$ are referred to as the "learning set." The learning set is partitioned into *V* sets of approximately size *n/V*. For each of the *V* folds, the "training set" consists of *V-1* sets, while the "test set" or "validation set" consists of the remaining 1 set. The observations in the training set are used to "train" the candidate algorithms and build candidate estimators. The test set is the used to evaluate the risk of each candidate algorithm, rotating *V* times in order to use each set as a test set once (85).

The motivation for Super Learner sprung from the age-old practice of analysts fitting various parametric model specifications (i.e., algorithms) according to preference and habit. For example, one might try a logistic regression with and without covariates, assessing the importance of each variable with p-values, keeping those with $p<0.20$, and finalizing the model based on the significance of each "confounder." This technique often results in biased estimates. Data-adaptive approaches avoid problems of analyst-induced bias by fitting several algorithms (e.g., a logistic regression) and using cross-validation to compare them in a supervised way.

Super Learner can also examine combinations of multiple algorithms that might "outperform" a single algorithm. "Algorithm" refers to any mapping from the observed data to the predictor, including both simple methods such as logistic regressions and complex approaches such as D/S/A. Importantly, the choice of algorithm (e.g., logistic regression with interaction terms) is no longer left to the analyst; instead, SL chooses the "best" algorithm based on minimizing the risk via cross-validation. This process optimizes the bias-variance tradeoff.

We can expect that a weighted average of some combination of algorithms would perform better (in terms of the loss function) than one algorithm alone. We can thus map the collection of candidate algorithms into a collection of their weighted averages, where each weighted average is its own candidate algorithm in the algorithm library. Cross-validation is ultimately used to select the optimal algorithm from the set of candidate algorithms. The weighted combination is part of the collection and indexed by a weight vector α. All α vectors sum to one, with each weight ≥ 0. The goal is to determine the combination that minimizes the cross-validated risk over the collection of weighted combinations. The weights that minimize the cross-validated risk are calculated from the regression of outcomes *Y* on the predicted values of the algorithms *Z*, using the library supplied by the analyst. The "better" algorithms will predict outcomes closest to those observed, and will thus be assigned larger coefficients in the regression and larger weights in the final prediction model.

The "oracle selector" is the considered the best estimator of the *K* algorithms in the Super Learner library; the oracle selector picks the algorithm minimizing the expected loss (risk) under the observed data's true distribution P_0. If all weighted combinations are used as the library, then the oracle selector chooses among this family of weighted

combinations. Van der Laan et al. have shown that the Super Learner performs as well as the oracle selector up to second order terms (85).

Super Learner process

First the data are input along with a collection of algorithms specified by the analyst (e.g., logistic regressions, classification and regression trees, D/S/A, etc.). The dataset is then divided into V folds of mutually exclusive sets. The first algorithm ($A1$) is fit on the first training set ($T1$). The validation set ($V1$) is then used to obtain predicted values ($Z_{A1, V1}$). The risk is then calculated for the validation set 1 ($V1$) using algorithm 1 ($A1$) and the predicted values $Z_{A1, T1}$. These steps are repeated for each algorithm contained in the library, such that after each algorithm has been run on the first fold, we have obtained predicted values Z for each algorithm, as well as an estimate of the risk for that particular validation set for each of the algorithms. This cycle is repeated for each of the V folds, such that each set is part of both the training and test/validation set for each algorithm.

At this point we have calculated predicted values for all participants for each algorithm, as well as corresponding risks within each test set for each algorithm. This means we have V risks estimated for each algorithm. The next step involves proposing a collection of weighted combinations indexed by the weight vector α. The predicted values Z are used to predict the outcome Y and estimate the vector α, minimizing the cross-validated risk over the collection of weighted combinations. The chosen combination equates to a new estimator which we will use to estimate predicted values with the observed data. The final step involves fitting each of the candidate algorithms on the entire observed data. The fits are combined with the weights obtained, generating the Super Learner prediction function, or the weighted combination of the candidate algorithms fit to the observed data.

The data-adaptive property automates the choice of algorithm, defining a loss function presents a performance measure, and cross-validation rigorously estimates the expectation of the loss function ("risk"). In addition, machine learning, or data-adaptivity, respects the statistical model. The combination of these properties makes Super Learner a rigorous, unbiased approach to modeling (85).

Targeted Maximum Likelihood Estimation

Once we have obtained $Q_0 = E_0(Y|A,W)$ from Super Learner, we can then update the initial estimate Q_0 by "targeting" it to optimize the bias-variance tradeoff. Data-adaptive estimation, such as Super Learner, optimizes the bias-variance tradeoff for $E_0(Y|A,W)$, which is not the best bias-variance tradeoff for our target parameter $\Psi(P_0)$ because the estimator of $E_0(Y|A,W)$ attempts to best predict average outcomes within every stratum defined by covariates W. Our collection of covariates W might be highly dimensional, yielding many strata and potentially causing the optimal estimator to sacrifice much bias in order to minimize variation across strata. Because we are interested in effects averaged across strata of W, we must first reduce the bias in the initial estimate of $E_0(Y|A,W)$, otherwise our effect estimate will be too biased. The optimal estimator of our target parameter must accept less bias to achieve the best tradeoff. Thus reducing the

bias in the initial estimate of $E_0(Y|A,W)$ to target our parameter of interest is necessary (85).

tMLE process

We first estimate $E_0(Y|A,W)$. This can be done with machine learning like Super Learner, which gives the optimal estimate. We then estimate predicted values of $Y|A,W$ for each subject, yielding $Q_n(A_i,W_i)$. Next, we modify the initial estimate of $Q_{0bar} = E_0(Y|A,W)$ by estimating the treatment mechanism $g_0(a|W) \equiv P_0(A = a|W)$ for $a \in A$ with Super Learner. The estimated treatment mechanism is used to compute a "clever covariate," such that the clever covariate for each subject i can be written as

$$H_n^*(A_i,W_i) \equiv [(I(A_i = 1)/g_n(1|W_i)) - (I(A_i = 0)/g_n(0|W_i))]$$

The initial estimate of $Q_{0bar} = E_0(Y|A,W)$ is then updated to yield

$$Q_{barn}^1(A,W) = expit(logit(Q_{barn}^0(A,W)) + \varepsilon_n H_n^*(A,W))$$

where ε_n is estimated by regressing Y on $H_n^*(A,W)$ using $logit(Q_{barn}^0(A,W)$ as an offset. This step "targets" the estimator such that it is a less biased estimate of the target parameter

$$\Psi(P_0) = E_W(E_0(Y|A=1,W) - E_0(Y|A=0,W)).$$

We then estimate the predicted outcomes for each subject under each possible treatment using the updated estimate

$$Q_{barn}^1(A,W) = expit(logit(Q_{barn}^0(A,W)) + \varepsilon_n H_n^*(A,W))$$

This involves setting a equal to each possible treatment level for each subject and obtaining corresponding outcomes for each treatment level, e.g.,

$$Q_{barn}^1(0,W) = expit(logit(Q_{barn}^0(0,W)) + \varepsilon_n H_n^*(0,W))$$

This step necessitates estimation of $H_n(a,W)$ and $logit(Q_{barn}^0(A,W))$ for each value of a. Finally, we estimate the target parameter $\Psi(P_0)$ as an empirical mean difference of predicted values for each level of treatment: $\Psi(\hat{P}_n) = 1/n \Sigma[Q_n(1,W_i) - Q_n(0,W_i)]$ (for binary treatment). We can also test the null hypothesis of $H_0 : \Psi_0 = 0$ using the test statistic $T = \Psi_n/\sqrt{(\sigma^2/n)}$ and the corresponding p-values calculated as $2\Phi(-|T|)$, where Φ is the cumulative distribution function of the standard normal distribution (85).

tMLE vs. other methods

Super learner trades off bias and variance with respect to Q_0, so although it estimates Q_0 well, it does not necessarily provide an unbiased estimate for Ψ_0. For that an additional tMLE step is necessary. Thus, tMLE offers major advantages in bias reduction over both traditional approaches such as linear or logistic regressions and more sophisticated approaches.

Targeted maximum likelihood estimation also offers advantages over models such as marginal structural models with inverse probability of treatment weighting (IPTW) and other substitution estimators like g-computation (note that tMLE is itself a substitution estimator). Given the distribution of the observed data, $P_0(O) = P_0(W,A,Y) = P_0(Y|A,W)$ $P_0(W)P_0(A|W)$, IPTW estimation involves estimating the treatment mechanism g0 (A|W), which only depends on one part of P_0, $P_0(A|W)$.
This means that the consistency of IPTW estimators depends on consistent estimation of $g_0(A|W)$. Similarly, "g-computation" estimators are functions of only $P_0(Y|A,W)$ and

$P_0(W)$ because they involve estimating $E_0(Y|A,W)$. Thus the consistency of the g-computation estimator depends on consistent estimation of $E_0(Y|A,W)$ (85).

One advantage of targeted maximum likelihood over IPTW and g-computation is its "double robustness." Because implementing tMLE depends on estimating both $E_0(Y|A,W)$ and the treatment mechanism $g_0(A|W)$, tMLE is consistent if either $E_0(Y|A,W)$ or $g_0(A|W)$ is estimated consistently. Furthermore, estimating both $E_0(Y|A,W)$ and $g_0(A|W)$ consistently results in an efficient estimator, and possibly the lowest asymptotic variance of any estimator. This property corresponds to lower finite sample bias and variance.

Proper inference and bias convergence depend on asymptotic linearity, a property that ensures convergence of the estimator g_n to a distribution satisfying the positivity assumption. Data-adaptive estimation can produce asymptotic linearity and improved efficiency with fewer assumptions than substitution estimators.

Targeted maximum likelihood estimators are substitution estimators involving generation of predicted values for $Y|A, W$ and setting treatment equal to each possible level. The differences in the predicted values across treatment levels then averaged, resulting in an average treatment effect. If the positivity assumption is violated, or nearly violated, meaning that not all covariate/exposure strata are represented, the IPTW estimator can be inconsistent and inefficient because it does not extrapolate over areas of sparse data. On the other hand, substitution estimators do extrapolate. Furthermore, substitution estimators ensure that the parameter estimate falls within its allowable range; substitution estimators respect the parameter space and are generally more robust to sparse data. However, the "curse of dimensionality" can pose estimation problems since the number of covariates available for generally small sample sizes precludes traditional non-parametric approaches. Targeted maximum likelihood estimation handles high-dimensional data very well (85).

Super Learner and Targeted Maximum Likelihood Estimation for direct effects

When assessing direct and indirect effects, we can define the intervention as A and the intermediate as Z. The g-computation formula for a direct effect of a treatment A on Y controlling for intermediate Z assumes the data structure $(W,A=(A,Z),Y)$, where (A,Z) can be thought of as a joint treatment. The direct effect when setting Z to a specified level $Z=z$ is calculated as $DE(z)=E_w\, E(Y|A=1,Z=z,W)\, E(Y|A=0,Z=z,W)$.

I will use SL to estimate the conditional mean of $Y|A,Z,W$, and implement tMLE to update of this initial estimator targeted towards the direct effect $DE(Z)$. The tMLE update requires fitting the conditional distribution $P(A=1|W)$ of $A|W$, as well as the conditional distribution $P(Z=z|A,W)$ of the intermediate $Z|A,W$. I will then use these estimates to construct a clever covariate $I(A=1,Z=z)/g(A,Z|W)-(A=0,Z=z)/g(A,Z|W)$, add it to initial regression of Y, on W,A,Z, and fit the coefficient epsilon. This updated regression will then be mapped into estimate $E_W(E(Y|A=1,Z=z,W)-E(Y|A=0,Z=z,W))$ for a fixed Z. This is a controlled direct effect (85).

tMLE for direct effect estimation

We assume that $(A,Z)\perp Y(a,z)|W$ for all possible a,z values of (A,Z). As for the estimation of controlled direct effects, this identifiability assumption is necessary for correctly estimating the natural direct effect:

$$\Psi^F(P_{U,X}) = E_W \Sigma E_0(Y(1, z) - Y(0, z)|W) g_0(z|W, A = 0)$$

$Y(1, z), Y(0, z)$ denote the counterfactual outcomes corresponding with the intervention $A=a, Z = z,$ and $g_0(z|W, 0) = P(Z = z|W, A = 0)$ (van der Laan, 2010, personal communication). If we further assume that $E(Y(1, z) - Y(0, z)|Z(0) = z, W) = E(Y(1, z) - Y(0, z)|W)$ (i.e., no confounding of the intermediate~covariates relationship), then

$$\Psi^F(P_{U,X}) = E(Y(1, Z(0)) - Y(0, Z(0)))$$

$Z(0)$ denotes the counterfactual intermediate variable when treatment $A = 0$. Without the additional assumption regarding no confounding of the intermediate~covariate relationship, $\Psi^F(P_{U,X})$ estimates the average controlled direct effect, which can be interpreted similarly as the natural direct effect (i.e., the direct effect of treatment averaged across levels of the intermediate). Therefore, the confounding assumption is not considered crucial. Thus the identifiability result for $\Psi^F(P_{U,X})$ amounts to no unmeasured confounding of the effect of the treatment on the mediator (i.e., $A \perp\!\!\!\perp Y_{az} \mid W$) and no unmeasured confounding of the effect of the mediator on the outcome (i.e., $Z \perp\!\!\!\perp Y_{az} \mid W$).

Because $g(A|W)$ will be correctly specified when treatment is randomized, then $\Psi_0^* = \Psi_0$, yielding an unbiased estimate of the natural direct effect. This result illustrates double-robustness, a crucial feature of tMLE. Importantly, estimators of Ψ_0 that solve the "efficient influence curve equation" are doubly robust to misspecification of Q_0 and g_0. In the case of randomized control trial data, we know $g_0(A|W)$, so our estimated will be unbiased if either $Q=Q_0$ or $g_Z=g_{0Z}$ (85).

I will use substitution estimators to estimate $\Psi(P_0)$, which necessitates estimation of the following:

- $Q_0 = E(Y|W,A,Z)$
- $g_0(z|W,A=0)$
- the marginal distribution of W

The marginal distribution of W is estimated by the empirical distribution of covariates $W_1, \ldots, W_n$. The conditional mean Q_0 is estimated with Super Learner. The conditional distribution g_Z of $Z|W,A=0$ can be estimated empirically as well. I will then follow the tMLE protocol of creating a fluctuation function of P_{hat0} such that its score at $\varepsilon = 0$ equals the efficient influence curve. We first consider the clever covariate

$$C_Y(g_0)(W,A,Z) = \{I(A=1)/g_0(1|W) * g_0(Z|W,A=0)/g_0(Z|W,A=1) - I(A=0)/g_0(0|W)$$

If g_n is an estimator of g_0, we denote $C_Y(g_n)$ be the estimator of this clever covariate.

Because Y is continuous and known to be bounded between (0.1), we define the fluctuation function as

$$logitQ_n^0(\varepsilon) = logit\ Q_n^0 + \varepsilon\ C_Y(g_n^0)$$

This logistic fluctuation function ensures that predicted values will fall between the bounds (0,1), subsequently bolstering the robustness of tMLE.

We then define the loss function $L(Q)$ as the minus-log(likelihood), such that $L(Q) = YlogQ = (1-Y)log(1-Q)$. Our goal is to minimize the empirical mean of loss along this parametric fluctuation model. $L(Q)$ is minimized when the coefficient of the clever covariate is zero, i.e., ε=0. The fluctuation amount (ε) is then estimated by minimizing the empirical mean of the loss function, yielding

$$\varepsilon_n^0 = argmin P_n L(Q_n^0(\varepsilon))$$

The clever covariate ε_n minimizes this function such that the efficient influence curve is solved and the estimate converges as desired in this one targeted MLE step and $Q_n^1 = Q_n^0(\varepsilon_n^0)$ and $P_n f$ is the sample mean of $f(O_i)$.

The next steps involve targeted estimation of g_{Z0}. Suppose Z has K+1 levels. Let $Z(k) = I(Z = k)$, k = 0, . . . ,K−1, such that Z=($Z(k)$:k = 0, . . . ,K−1). Let $g_Z(k) = P(Z(k) \mid Z(0), . . . ,Z(k-1),W,A)$, such that $g_Z = \Pi\, g_Z(k)$. Let g_{Zn} denote the estimator corresponding with $g_{Z(k),n}$, k = 0, . . . ,K − 1. We then fluctuate

$$logit\, g_{Z(k),n}(\varepsilon)(1|W,A) = logit\, g_{Z(k),n}(1|W,A) + \varepsilon C_{Z(k)}(g_n^0, Q_{bar1})(W,A)$$

with the loss function

$$L(g) = \Sigma[Z(k)\, log\, g_{Z(k)}(1|W,A) + (1-Z(k))\, log\, g_{Z(k)}(0|W,A)]$$

and we will estimate ε with maximum likelihood estimation such that

$$\varepsilon_{ng}^0 = argmax P_n L(g_Z(\varepsilon))$$

and $g_n^1 = g_n^0(\varepsilon_{ng}^0)$ is the update. The final step is to determine the clever covariate $C_{Z(k)}$ which minimizes the loss function.

We have now defined the updating process $Q_n^1 = Q_n^0(\varepsilon_n^0)$, given Q_n^0, g_n^0, $g_n^1 = g_n^0(\varepsilon_{ng}^0)$, given Q_n^1, g_n^0. Finally, we evaluate $\psi_n^* = \Psi(Q_n^1, g_n^1)$ which solves the efficient influence curve equation and is this doubly robust and efficient under regularity conditions {van der Laan, 2011}.

V. STATISTICAL ANALYSES

Additive treatment effects and controlled direct effects (Aims 1 and 2)

The targeted maximum likelihood estimation package in R was implemented using Super Learner to predict *Q*. The following algorithms were included for *Q* in the Super Learner library based on conversations with Dr. van der Laan: "SL.gam", "SL.gam.3", "SL.gam.4", "SL.gbm.1", "SL.gbm.2", "SL.glm", "SL.glmnet", "SL.glmnet.alpha25", "SL.glmnet.alpha50","SL.glmnet.alpha75", "SL.polymars", "SL.randomForest", "SL.ridge", "SL.svm", "SL.bayesglm", "SL.step", "SL.step.interaction", "SL.bart", "SL.myglm2", "SL.myglm2.cell", "SL.myglm2.pda", "SL.myglm.cellbypda2", "SL.myglm3", "SL.DSA". The goal was to include algorithms with two- and three-way interaction terms (e.g, "SL.myglm2", "SL.myglm.cellbypda2"- these are wrappers written to force inclusion of interaction variables) as well as data-adaptive algorithms like DSA. The following libraries were included for *g*: "SL.glm","SL.step", and "SL.DSA2." Because data come from a randomized trial and the treatment mechanism *g* is thus known, we felt comfortable using a smaller algorithm library for estimating *g.* Note that various libraries were implemented for comparison's sake. Including various subsets of these algorithms did not change results.

The data were analyzed excluding those participants missing *Y* (percent days abstinent after 16 weeks). This was for consistency with the COMBINE publication of trial results in JAMA (2006). Note that results were not sensitive to this exclusion. The treatment variable *A* = (placebo, naltrexone groups, CBI groups, naltrexone + CBI groups) with placebo as the reference group. The set of covariates *W* = (baseline percent days abstinent, site of treatment administration) was also based on what the COMBINE investigators had done. For controlled direct effect estimation, *Z* = (cravings) dichotomized into "high" vs. "low" based on the median split; this decision was both practical (the tMLE package can only handle two levels of *Z*) and theoretical (interest in high vs. low cravings). For controlled direct effect estimation, participants missing *Z* were excluded in addition to those missing Y. The dichotomization procedure was analogous when examining stress as an intermediate.

The estimation process yields additive treatment effects, controlled direct effects at each of the two levels of *Z*, as well as the variances, p-values, and confidence intervals for each estimate. Inference was based on the 95% confidence intervals.

Natural direct effect and indirect effect analyses (Aims 3 and 4)

The current R software does not compute natural direct effects (NDE). A biostatistics student, Sam Lendle, was recruited to write the NDE code. The code ultimately estimates $E(Y|A=1, Z, W)$ and $E(Y|A=0, Z, W)$ for everyone and averages the estimates over the distribution of *W* and *Z* for those who are not treated (i.e., averages $E(Y|A=1, Z, W)$ and $E(Y|A=0, Z, W)$ over $W, Z \mid A= 0$). Specifically, Super Learner is used to predict $Y \mid W, Z$ while plugging in $A = 0$ and $A = 1$ for everyone, and yields an estimate of $\{E(Y|A=1, Z, W)-E(Y|A=0, Z, W)\}$. This approach to NDE estimation mimics that for the average treatment effect among the treated (85). Here is the distribution of *W,*

$Z \mid A = 0$ (i.e., the distribution of covariates and the intermediate in the untreated population) is used because the natural direct effect assumes an untreated population.

The NDE code yields estimates for the natural direct effect and indirect effect as well as the variances for each. Ninety-five percent confidence intervals were calculated using the standard normal formula. Inference was based on the 95% confidence intervals.

VI. RESULTS

Additive treatment effects

Among those not missing *Y*, additive treatment effects are similar in pattern but stronger than magnitude compared to results published in *JAMA* (Table 1). Compared to placebo, the COMBINE study investigators report differences in average 16-week PDA of 5.5, 4.1, and 2.0, for naltrexone, CBI only, and naltrexone + CBI, respectively; these differences are all reported as significant at the $p<0.05$ level. Because the authors used mixed models, these results should be interpreted as a subject's average PDA over time when given treatment vs. placebo, conditional on treatment site and baseline PDA (Hubbard, to GEE). Results from the tMLE package show differences in average 16-week PDA of 6.3, 5.7, and 4.6, for naltrexone, CBI only, and naltrexone + CBI, respectively. Results when excluding those missing *Z* data are slightly higher and the same in pattern for naltrexone and the Combination but slightly different for CBI, whose effect diminishes and is no longer significant. All effects are significant except for CBI when excluding those missing both Y and Z. The pattern is similar to the *JAMA* results, as naltrexone only appears to have the strongest effect, followed by CBI only, then the naltrexone + CBI combination.

Table 1: Additive Treatment Effects

Sample	**Naltrexone RD[1] (SD)**	**CBI RD (SD)**	**Combo RD (SD)**
JAMA results	5.5 (nr[2])	4.1 (nr)	2.0 (nr)
Excluding those missing Y	6.3 (2.3)**	5.7 (2.2)**	4.6 (2.3)*
Excluding those missing Y and Z	7.0 (2.4)**	3.9 (2.4)	5.3 (2.5)*
Entire baseline sample	6.3 (2.2)**	5.8 (2.2)**	4.6 (2.3)*

[1] Risk Difference
[2] Not reported in JAMA paper
*p<0.05
**p<0.01

Controlled direct effects

Table 2 displays controlled direct effect results across 4- and 12-week high vs. low craving levels. Effects stratified by cravings measured at 4 weeks show that when cravings are low at 4 weeks, none of the active treatments appear to be effective relative to placebo. However, when cravings are high at 4 weeks, all of the treatments are similarly effective. For example, if cravings are high for everyone at 4 weeks and everyone gets CBI only, average PDA is 12.8 percentage points higher than if cravings are high for everyone at 4 weeks and everyone got placebo. Results are similar for naltrexone groups and even stronger for the Combination. The differences in significant effects for high vs. low cravings suggest that treatment effects may be modified by 4-week cravings' levels.

Controlled direct effects stratified by cravings measured at 12 weeks show that when cravings are low at 12 weeks, again none of the active treatments are effective compared to placebo. When cravings are high at 12 weeks, the naltrexone groups and combo groups show effects, while the CBI groups do not; for example, if cravings are high for everyone at 12 weeks and everyone gets naltrexone only, average PDA is 10.7 percentage points higher than if cravings are high for everyone at 12 weeks and everyone got placebo. Results are similar for the Combination groups and null for CBI groups.

Table 2: Controlled direct effects across craving levels

		Naltrexone RD (SD)	CBI RD (SD)	Combo RD (SD)
4-week Cravings	Low	2.1 (3.1)	1.8 (3.0)	3.3 (3.1)
	High	12.5 (4.0)**	12.8 (3.8)*	16.0 (3.9)**
12-week Cravings	Low	1.4 (2.4)	-0.04 (2.2)	1.9 (2.2)
	High	10.7 (3.7)**	6.5 (3.6)†	9.1 (3.9)*

†p<0.10
*p<0.05
**p<0.01

Table 3 displays controlled direct effect results across 4- and 12-week high vs. low stress levels. Controlled direct effects stratified by stress measured at 4 weeks show that when stress is low at 4 weeks, none of the active treatments appear to be effective relative to placebo. CBI and the Combination appear to have borderline significant effects, yet the point estimates are lower than those for high 4-week stress.

When stress is high at 4 weeks, strong effects obtain in the naltrexone groups and the Combination groups. For example, if stress is high for everyone at 4 weeks and everyone gets naltrexone, average PDA is 9.6 percentage points higher than if stress is high for everyone at 4 weeks and everyone got placebo. Results are stronger for the Combination, with a 10.5 percentage point difference in PDA relative to placebo when stress is high at 4 weeks. Thus stress appears to moderate the naltrexone and combination group effect. Effects in the CBI group are borderline significant, yet so similar to those when stress is low that a meaningful treatment*stress interaction seems unlikely for the CBI groups.

The pattern of the controlled direct effects stratified by stress measured at 12 weeks us similar to those stratified by 4-week stress. Again, both naltrexone and the Combination are markedly more effective when stress is high, while CBI effects do not appear to change across stress levels. For example, if stress is high for everyone at 12 weeks and everyone gets the Combination, average PDA is 17.0 percentage points higher than if stress is high for everyone at 12 weeks and everyone got placebo. Results are slightly weaker for the naltrexone groups and null for CBI groups.

Table 3: Controlled direct effects across stress levels

		Naltrexone RD (SD)	CBI RD (SD)	Combo RD (SD)
4-week Stress	Low	3.2 (3.7)	5.9 (3.5) †	6.1 (3.5) †
	High	9.6 (4.0)*	6.7 (3.9) †	10.5 (3.9)*
12-week Stress	Low	5.8 (3.0) †	5.1 (3.0)	3.8 (3.2)
	High	11.4 (4.8)*	6.9 (4.5)	17.0 (4.2)**

†p<0.10
*p<0.05
**p<0.01

Natural direct effect and indirect effect analyses

All analyses were performed for the sample of those not missing *Y* or *Z* = (*4-week cravings, 12-week cravings, 4-week stress, 12-week stress*), controlling for baseline measures of PDA, cravings, and stress, as well as treatment site. Table 4 shows which treatment affected cravings and stress at each timepoint; tMLE was implemented for these estimations, where each mediator (e.g., cravings at 4 weeks) was modeled as the dependent variable. Table 5 displays natural direct and indirect effect results via cravings at each timepoint. Table 6 displays natural direct and indirect effect results via stress at each timepoint. Table 7 displays natural direct and indirect effect results for the set of (*4-week cravings, 12-week cravings, 4-week stress, 12-week stress*). Cravings and stress at each timepoint (4 weeks, 12 weeks) were examined as mediators separately and together in order to gauge how each potential mediator contributed to indirect effects.

Treatment effects on mediators

As an intermediate step to satisfy curiosity, cravings and stress at each timepoint were analyzed as dependent variables prior to natural direct effect analyses to assess whether each treatment was affecting the proposed intermediates. Targeted maximum likelihood estimation for average treatment effects was implemented for each proposed intermediate (Table 4, column 1) where the intermediate was entered as *Y*. Table 4 shows which treatments predicted which mediators at each timepoint. Parameter estimates are not shown because this analysis was just to make sure that the treatments were affecting the mediators as hypothesized and actual effect sizes were not of substantive interest. Naltrexone appears to work on both earlier and later cravings. CBI and the Combination work on later cravings. No treatment affects stress at either 4- or 12-weeks.

Table 4: Do treatments predict mediators?

	Naltrexone RD (SD)	CBI RD (SD)	Combo RD (SD)
4-week Cravings	Yes, reduces (p=0.005)	No	No
12-week Cravings	Yes, reduces (p=0.004)	Yes, reduces (p=0.02)	Yes, reduces (p=0.04)
4-week Stress	No	No	No
12-week Stress	No	No	No

Cravings

First, 4-week cravings were examined (Table 5). Naltrexone and the Combination effects appeared to be mediated, at least in part, by 4-week cravings, as the indirect effects for each are significant. The indirect effect for CBI is not significant, indicating that 4-weeks cravings do not mediate CBI effects. These results are all in line with those shown in Table 4.

When examining 12-week cravings, the indirect effects for all three treatments are significant; thus each treatment appears to affect 12-week cravings such that a 3.5 to 4.9-percentage point increase in PDA results. All treatment effects are at least partially mediated by 12-week cravings. These results are again consistent with those shown in Table 4.

When analyzing both 4- and 12-week cravings simultaneously, results for naltrexone and the Combination are similar to those assessing only 12-week cravings, possibly indicating a threshold effect of cravings for these treatments: craving reduction at one timepoint yields a 4-5 percentage-point increase in PDA, and accounting for craving reduction at other timepoints does not strengthen this effect.

On the other hand, the indirect effect for CBI is both smaller in magnitude compared to the other treatments, and no longer significant when accounting for both 4- and 12-week cravings. The direct and indirect effects when controlling for both 4- and 12-week cravings amount to averaging across the effects for 4- and 12-week cravings alone; this could explain the disappearance of significant effects for the CBI group.

Interestingly, the direct effects of the Combination remain significant at each timepoint, implying that this treatment may work through additional mechanisms other than cravings. In contrast, the direct effects for naltrexone at each timepoint are not significant, suggesting that craving reduction is the predominant mediating variable for these treatments. The results for CBI are mixed: although CBI does appear to work through 12-week cravings, the result disappears when adding 4-week cravings to the model. Yet because CBI works through cravings at 12 weeks, and given that it takes longer to have an effect (as shown in Table 4), averaging across 4- and 12-week cravings is probably not useful for this group anyway.

To quantify the proportion of treatment effects attributable to cravings, the following formula was used: (indirect effect)/(indirect effect + direct effect)*100. Thus 47-63% of each treatment effect can be attributed to craving reduction.

Table 5: Natural direct effects across craving levels

		Naltrexone RD (SD)	CBI RD (SD)	Combo RD (SD)
4-week Cravings	Direct Effect	3.3 (2.4)	5.7 (2.3)*	6.7 (2.3)**
	Indirect Effect	3.9 (1.2)**	-0.001 (1.2)	1.4 (1.1)*
12-week Cravings	Direct Effect	4.1 (2.2) †	2.3 (2.0)	5.4 (2.0)*
	Indirect Effect	4.9 (1.7)*	3.9 (1.7)*	3.5 (1.5)*
4- & 12-week Cravings	Direct Effect	3.7 (2.1) †	3.0 (1.9)	4.7 (1.9)*
	Indirect Effect	4.2 (1.7)*	2.8 (1.8)	4.3 (1.8)*
% effect explained by 4- & 12-week Cravings		53.2%	62.9% explained by 12-week Cravings	47.8%

†p<0.10
*p<0.05
**p<0.01

Stress

Although no treatment appeared to affect stress, natural direct effects were still analyzed across stress levels analogously to cravings NDE assessment. First, 4-week stress was examined. All indirect effects were non-significant while all direct effects remained strong and significant. Although there is some evidence of mediation by 12-weeks stress for the naltrexone group, as shown by the significant indirect effect, results were similar for 12-week stress, as well as for both 4- and 12-week stress when analyzed simultaneously. This pattern of results was expected given that no treatment affects stress levels at either timepoint (see Table 4), and thus the proportion explained by stress was not calculated (see Table 6).

Table 6: Natural direct effects across stress levels

		Naltrexone RD (SD)	CBI RD (SD)	Combo RD (SD)
4-week Stress	Direct Effect	6.5 (2.7)*	6.4 (2.7)*	7.5 (2.7)**
	Indirect Effect	-0.001 (0.40)	-0.89 (0.46) †	0.25 (0.41)
12-week Stress	Direct Effect	6.1 (2.7)*	6.2 (2.6)*	8.5 (2.7)**
	Indirect Effect	2.3 (0.77)**	-0.18 (0.68)	0.02 (0.53)
4- & 12-week Stress	Direct Effect	7.7 (2.6)**	5.7 (2.6)*	8.4 (2.8)**
	Indirect Effect	0.15 (0.87)	-0.11 (0.74)	0.31 (0.54)
% effect explained by 4- & 12-week Stress		n/a	n/a	n/a

†p<0.10
*p<0.05
**p<0.01

Cravings and stress

Finally, 4- and 12-week cravings and 4- and 12-week stress were analyzed simultaneously. Because no treatment appeared to affect stress at either timepoint, inclusion of stress in the multiple mediator model was not expected to change results. Table 7 shows that analyzing all mediators, i.e., cravings at 4 and 12 weeks and stress at 4 and 12 weeks, yields results similar to those for analyzing simply cravings at 4 and 12 weeks. Thus stress does not appear to explain any treatment effects, as expected given that no treatment predicted intermediate stress levels. Cravings and stress combined at 4 and 12 weeks appears to explain about 50% of all treatment effects.

Table 7: Natural direct effects across both craving and stress levels

		Naltrexone RD (SD)	CBI RD (SD)	Combo RD (SD)
4- &12-week Cravings & Stress	Direct Effect	3.8 (2.1)†	2.8 (1.8)	4.6 (2.1)*
	Indirect Effect	4.1 (1.7)*	3.0 (1.8)†	4.1 (1.7)*
% effect explained by 4- & 12-week Cravings & Stress		51.9%	51.7%	47.1%

†p<0.10
*p<0.05
**p<0.01

VII. DISCUSSION

Brief summary of results

Controlled direct effect analyses show that naltrexone, CBI, and the Combination all appear to work best when cravings are high, while none of the treatments work when cravings are low; these results imply that cravings levels, at 4 and 12 weeks, moderate treatment effects. The differences in significant effects for high vs. low cravings and between the 4-week vs. 12-week timepoints suggest that while CBI and the Combination may be more effective for those with high cravings at 4 weeks, cravings at 12 weeks do not appear to modify CBI treatment effects nearly as strongly. In contrast, naltrexone appears effective for those with high cravings regardless of when cravings occur. Similarly, naltrexone and the Combination work better when stress is high, again for both 4- and 12-week stress levels. CBI effects do not appear moderated by stress levels, regardless of timing.

Natural direct effect analyses show that all treatment effects are at least partially mediated by cravings, with craving reduction explaining 47-63% of treatment effects. When assessing treatment effects on cravings, naltrexone affects cravings earlier (4 weeks), while CBI works later (12 weeks). Unexpectedly, the Combination does not appear to affect 4-week cravings, and only has a borderline association with 12-week cravings; perhaps naltrexone's effects on craving disappear when combined with CBI. Also contrary to hypotheses, no treatment appears to affect stress levels at either 4 or 12 weeks.

Consistent with the above results, neither stress nor cravings appear to fully explain CBI's or the Combination's effect, as evidenced by the significant natural direct effects. Unlike the naltrexone condition, the persistence of a significant natural direct effect for the Combination implies that this particular treatment may have important mediators besides cravings or stress.

Results suggest that naltrexone and CBI operate to some degree through the same mediator, but at different times, such that once cravings have been reduced by the naltrexone, the behavioral intervention can no longer affect cravings. Controlled direct effect analyses show that none of the treatments work when cravings are low, while the pre-NDE analyses show that naltrexone affects earlier cravings while CBI works later. Thus, hypothetically, a patient taking both naltrexone and CBI might experience craving reduction early on (since naltrexone works on 4-week cravings) but then receive no benefit from CBI (because no treatments work when cravings are low).

This set of results has two important implications. First, if craving reduction is the central mediator for CBI (as suggested by natural direct effect results at 12 weeks), and CBI doesn't work when cravings are low (as evidenced by controlled direct effect results), then reducing cravings past a certain point seems moot; i.e., the reason that CBI doesn't work when cravings are low is because CBI works via craving reduction, and once one's cravings are low, lowering them more will not help outcomes. The second noteworthy implication of this set of results is the potential to explain why the Combination was not more effective that either monotherapy. Again, the findings suggest that supplementing naltrexone with CBI may be futile because the two treatments are working through the same mechanism, just at different times.

Connection with literature to date

The idea of a cravings' threshold has been explored, although not relative to naltrexone or behavioral therapies. A single-case report suggested that baclofen, a GABA derivative used primarily to treat spasticity, may completely eliminate cravings in alcohol-dependent patients, yet the body of research as a whole implies that no single neural pathway completely explains the occurrence of cravings (88). Thus more research is needed to understand how to completely suppress cravings. Furthermore, these results do not shed light on any threshold effects of cravings on drinking; in other words, the threshold of zero cravings makes theoretical sense and has been empirically shown to exist, yet no current data demonstrate whether reaching a particular point of cravings (e.g., "very few") is enough to predict abstinence.

Although COMBINE investigators did expect some distinct mechanisms of action for each condition (89), these results show that cravings may be the most salient mechanism for both the naltrexone and CBI conditions. Furthermore, there is research that supports that both treatments affect cravings, as naltrexone may reduce craving by blocking the conditioned activation of endogenous opioids (70) while CBI might alleviate craving via cognitive and behavioral coping skills, enhanced self-efficacy, or increased motivation for abstinence (23).

Results from numerous clinical trials have demonstrated that naltrexone reduces craving levels (24, 25, 90-92), and patients with the highest levels of baseline craving appear to benefit especially from naltrexone (24, 72, 93). Richardson et al. similarly reported greater naltrexone efficacy for patients with high cravings during the first six weeks of treatment. Heilig (2006) claimed that because endogenous opioids modulate the positive reinforcement of alcohol consumption, naltrexone would be expected to benefit subjects whose disease is mostly characterized by reward craving (5). Although types of cravings cannot be distinguished with the current data, these past findings are consistent with the results of our analyses.

Similarly, a trial examining the utility of matching patients to either naltrexone or placebo and either relapse prevention therapy or supportive therapy based on pretreatment craving, alcohol dependence severity, and cognitive measures of learning and memory showed that those with higher levels of craving and poorer cognitive functioning fared best on naltrexone compared to placebo (72). In terms of the behavioral therapy components, relapse prevention therapy was associated with worse outcomes for those with lower levels of verbal learning, but better outcomes for those with higher levels of verbal learning (72). Although the current results do not allow for comparisons regarding cognitive function, the current results do corroborate the findings regarding naltrexone effect modification by high cravings.

CBI's causal chain is less well understood than that of naltrexone. In addition, CBI was developed for COMBINE, precluding past examination of potential mediators. Although the biological mechanism is not known, CBI may also affect abstinence via craving reduction (71), which may function relative to cognitive and behavioral processes (23). Cravings could result from conditioned response to prior gratification or withdrawal symptoms. Ultimately, cognitive/emotional states, positive and negative cues, mental processes, mood, and self-efficacy all influence craving levels, and CBI address these aspects of recovery (23). Furthermore, coping and self-efficacy also predict abstinence;

CBI explicitly addresses these individual characteristics, implying that CBI may affect craving, either directly or indirectly via increased coping skills or self-efficacy. Others have claimed that the CBI focus on inter- and intrapersonal functioning highlights personal improvements in self-efficacy or mood as predictors of abstinence, perhaps highlighting important variables upstream from cravings (71).

Although stress moderated naltrexone and the Combination effects (most likely because the Combination included naltrexone), stress was not a strong mediator. Furthermore, stress did not appear to moderate the CBI effects. Some have asserted that naltrexone effects are mediated through the stress axis (57, 94), and stress reduction is one skill taught during CBI sessions (53). Here, natural direct effect results at 12 weeks suggest that stress reduction may be at play for naltrexone but not CBI.

Although no available studies have shown naltrexone differences across stress levels, a study among patients with PTSD showed both better drinking outcomes for active medication (naltrexone, disulfiram, or the Combination) than placebo, as well as overall improvement of PTSD symptoms (95). With the understanding that PTSD and stress as measured in COMBINE are not the same, these results suggest that naltrexone may alleviate symptoms, such as cravings, associated with extreme stress.

A trial led by Kranzler indicated a bimodal distribution of mood states preceding both any drinking and heavy drinking in that individuals drank more on days which the felt higher levels of positive or negative mood states (96). Kranzler also reported that naltrexone weakened positive associations between heavy drinking and both positive and negative mood, and that the desire to drink mediated the effects of positive mood on any drinking (96). Although the relationship is complex, mood and stress are thought to predict one another (97), so perhaps the desire to drink (i.e., craving) also mediates the effects of stress on drinking as well

Implications for treatment and policy

The most significant implication for treatment regards the potential futility of combining pharmaceuticals and behavioral therapies. A COMBINE Primary Investigator, Raymond Anton, has claimed, "Because CBT and naltrexone share common mechanisms of action, such as craving reduction and relapse prevention, these therapies are likely to be well suited to use in combination" (61). However, if naltrexone and CBI work through the same pathway, i.e., through craving reduction, and if a threshold effect of cravings exists, then combining the two would not improve outcomes. Thus prescribing both could waste time and money. Perhaps patients could choose between a pharmaceutical and behavioral therapy, as some have shown that patients tend to have better outcomes when allowed to choose their own treatments (98). However, there have been no preference trials for pharmaceutical vs. behavioral alcoholism treatments to date, and this remains a rich area for research. In addition, the Combination treatment may operate through different mediators, and understanding what these potential mechanisms are would further inform treatment prescription.

Craving reduction explains 47-63% of treatment effects, suggesting that reducing cravings should greatly improve abstinence. Health care providers could work with patients to learn how to best control cravings, whether through naltrexone or CBI-type methods. Furthermore, the controlled direct effects' results suggest that patients with high

cravings may benefit from all treatments, while those with high stress may benefit most from naltrexone.

Meyer (2000) has argued that craving reduction is key, and that understanding "situational craving," or alcoholics' responses to cues in their favorite drinking environments would enhance current treatment (99). Witkiewitz et al. used the COMBINE data to see how the cravings module specific to the CBI groups affected heavy drinking, and found that those who received the module had fewer heavy drinking days, and that receiving the module moderated the effects of negative mood on heavy drinking; these moderated effects were mediated by craving reduction (100). Others have shown that cravings, as measured by the Obsessive Compulsive Drinking Scale (also used in COMBINE), is a strong predictor of relapse among outpatients (101). The controlled direct effect results also imply that those with low cravings or low stress may not do any better on naltrexone, CBI, or the Combination than they would on placebo. This result suggests the importance of further research on optimal treatments for low-cravings or low-stress subgroups.

Stress reduction is a prominent reason for alcohol use (102), and thus deserves focus as an intermediate between treatment and outcomes. A review of pharmacological treatments for alcohol dependence concluded that receptors for stress-related neuropeptides, such as corticotropin-releasing factor (CRF), are important targets for drug development (5). A later review by Spanagel and Kiefer echoed this conclusion, claiming that neural CRF signals combined with the endocannabinoid system integrate stress-related events and thereby mediate drinking outcomes (21). Despite the current null results for stress as a mediator of the COMBINE effects, future analyses should examine how treatments work for those with high stress and whether treatments can (or should) be adjusted to address stress issues.

Directions for future research

First, the reasons why the Combination is not affecting cravings strongly at either timepoint need to be identified. Treatment adherence did not differ across conditions, and thus adherence cannot explain this unexpected finding. Similarly, understanding why CBI and the Combination are not affecting stress, when theoretically the CBI should reduce stress by introducing coping skills, is also crucial.

The treatments all work best for those with high cravings, most likely because they reduce cravings. But why do naltrexone and the Combination work best for those with high stress if neither actually affects stress levels? This question should be explored in future analyses.

The persistence of significant natural direct effects for the Combination implies that there are other important mechanisms at work for this condition besides cravings (and stress). Future mediation analyses should assess other potential moderators and mediators, such as social network composition or problem severity measures. Moderation of naltrexone by genetic factors may also be possible. Monterosso et al. (2001) has reported that patients with high familial loadings for alcohol problems may also benefit especially from naltrexone (73). However, Capone et al. reported that although family history of alcoholism (FHA) did significantly predict drinking outcomes within the COMBINE sample, FHA did not moderate naltrexone efficacy (103). Rohsenow et al. has also shown

that the percentage of relatives with problem drinking modify naltrexone effects in that those with higher percentages fared better on naltrexone (104). Krystal and others have also shown genetics factors to moderate naltrexone effects (90, 105, 106). Thus further exploration of treatment effect modification by genetic factors may be warranted.

Other moderators may be explored as well. For example, a recent efficacy trial showed that naltrexone increased PDA in all participants, reduced urge levels in participants with younger age of alcoholism onset, increased time between drinks in participants who had more relatives with alcohol problems, and decreased alcohol's stimulating effects in women. On the other hand, both Greenfield et al. and Kranzler et al. have claimed that naltrexone effects do not differ between genders (107, 108). Future analyses could include assessing whether age of onset, family history or gender modify treatment effects shown here.

In a separate trial, those with more antisocial traits had less heavy drinking on naltrexone relative to placebo, whereas patients low on antisocial traits did not benefit from naltrexone (104). A naltrexone/CBT combination trial led by Anton showed that motivated patients with moderate dependence did best on naltrexone combined with outpatient cognitive behavioral therapy, and that naltrexone increased control over urges to drink and thoughts about drinking among those patients (109). Anton has further argued that psychiatric conditions like depression or anxiety can affect cravings, and that stress can trigger relapse (110). Among patients diagnosed with depression, at least one study has shown that although disulfiram appeared to reduce cravings more than naltrexone, both medications were effective in reducing actual alcohol use (111). Witkiewitz et al. have also shown that cravings mediate the effects of depressive symptoms on substance use among patients receiving usual care, but not among those attending mindfulness-based relapse prevention sessions that focused on coping with emotional discomfort and lessening conditioned responses of craving; these results support a moderated mediation effect (112). Results from a small, uncontrolled pilot study also showed that learning to manage negative cognitions and affect, as well as acceptance-based strategies for preventing relapse may also reduce craving and subsequent relapse (113). Future analyses could also examine whether personality characteristics or disorders modify treatment effects.

Finally, the dependent variable examined here was percent days abstinent (PDA). Targeted maximum likelihood estimation could be expanded to include hazard models for evaluating how cravings and stress mediate other outcomes, such as time to next drinking day.

CONCLUSION

The results of this dissertation are important both methodologically and substantively. This work contributes to the mediation/direct effects methods' literature by discussing the strengths and weaknesses of traditional approaches, arguing for an alternate "causal inference" approach with both statistical and philosophical advantages, and applying the causal inference approach to a dataset from the COMBINE study, a large RCT which had previously yielded surprising results. Furthermore, tMLE, the statistical approach taken here, has not been applied to questions of direct/indirect effects in the literature to date.

Substantively, this dissertation explores the unexpected results from the COMBINE study and suggests that the lack of advantage of the Combination naltrexone + CBI therapy is due to naltrexone and CBI's working through the same pathway: cravings. These results imply that combining naltrexone and CBI may not be optimal because of their similar mechanisms and that, generally speaking, those providing healthcare recommendations should consider competing mechanisms when prescribing treatment.

REFERENCES

1. Kreek MJ, LaForge KS, Butelman E. Pharmacotherapy of addictions. *Nature Reviews Drug Discovery* 2002;1(9):710-26.
2. Fuller R, Branchey L, Brightwell D, Et Al. Disulfiram Treatment Of Alcoholism. *Jama-Journal of the American Medical Association* 1986;256(11):1449-55.
3. Lundwall L, Baekelan.F. Disulfiram Treatment Of Alcoholism - Review. *Journal of Nervous and Mental Disease* 1971;153(6):381-&.
4. Johnson B. Update on neuropharmacological treatments for alcoholism: Scientific basis and clinical findings. *Biochemical Pharmacology* 2008;75(1):34-56.
5. Heilig M, Egli M. Pharmacological treatment of alcohol dependence: Target symptoms and target mechanisms. *Pharmacology & Therapeutics* 2006;111(3):855-76.
6. Hughes J, Cook C. The efficacy of disulfiram: A review of outcome studies. *Addiction* 1997;92(4):381-95.
7. Paille F, Guelfi J, Perkins A, Et Al. Double-Blind Randomized Multicenter Trial Of Acamprosate In Maintaining Abstinence From Alcohol. *Alcohol and Alcoholism* 1995;30(2):239-47.
8. Kranzler H. Extended-release intramuscular naltrexone - A viewpoint. *Drugs* 2006;66(13):1753-4.
9. Lhuintre J, Moore N, Tran G, Et Al. Acamprosate Appears To Decrease Alcohol Intake In Weaned Alcoholics. *Alcohol and Alcoholism* 1990;25(6):613-22.
10. McEntee WJ, Crook TH. Glutamate - Its Role In Learning, Memory, And The Aging Brain. *Psychopharmacology* 1993;111(4):391-401.
11. Johnson B, Ait-Daoud N, Anton R, et al. Topiramate as treatment for alcohol dependence - Reply. *Jama-Journal of the American Medical Association* 2008;299(4):406-7.
12. Littleton J. Acamprosate In Alcohol Dependence - How Does It Work. *Addiction* 1995;90(9):1179-88.
13. Bisaga A, Popik P. In search of a new pharmacological treatment for drug and alcohol addiction: N-methyl-D-aspartate (NMDA) antagonists. *Drug and Alcohol Dependence* 2000;59(1):1-15.
14. Soyka M, Chick J. Use of acamprosate and opioid antagonists in the treatment of alcohol dependence: A European perspective. *American Journal on Addictions* 2003;12:S69-S80.
15. Kiefer F, Mann K. New achievements and pharmacotherapeutic approaches in the treatment of alcohol dependence. *European Journal of Pharmacology* 2005;526(1-3):163-71.
16. Gual A, Lehert P. Acamprosate during and after acute alcohol withdrawal: A double-blind placebo-controlled study in spain. *Alcohol and Alcoholism* 2001;36(5):413-8.
17. Kranzler HR, Van Kirk J. Efficacy of naltrexone and acamprosate for alcoholism treatment: A meta-analysis. *Alcoholism-Clinical and Experimental Research* 2001;25(9):1335-41.
18. Mann K. Pharmacotherapy of alcohol dependence - A review of the clinical data. *Cns Drugs* 2004;18(8):485-504.
19. Carmen B, Angeles M, Ana M, et al. Efficacy and safety of naltrexone and acamprosate in the treatment of alcohol dependence: a systematic review. *Addiction* 2004;99(7):811-28.

20. Chick J, Health Technology Board of S. A health technology assessment of prevention of relapse in alcohol dependence. *Alcoholism Clinical and Experimental Research* 2003;27(5 Supplement):112A.
21. Spanagel R, Kiefer F. Drugs for relapse prevention of alcoholism: ten years of progress. *Trends in Pharmacological Sciences* 2008;29(3):109-15.
22. Litten R, Allen J. Advances in development of medications for alcoholism treatment. *Psychopharmacology* 1998;139(1-2):20-33.
23. Addolorato G, Leggio L, Abenavoli L, et al. Neurobiochemical and clinical aspects of craving in alcohol addiction: A review. *Addictive Behaviors* 2005;30(6):1209-24.
24. Volpicelli J, Alterman A, Hayashida M, Et Al. Naltrexone In The Treatment Of Alcohol Dependence. *Archives of General Psychiatry* 1992;49(11):876-80.
25. O'Malley SS, Jaffe AJ, Chang G, et al. Naltrexone and coping skills therapy for alcohol dependence: A controlled study. *Archives of General Psychiatry* 1992;49(11):881-7.
26. O'Malley SS, Jaffe AJ, Chang G, et al. Six-month follow-up of naltrexone and psychotherapy for alcohol dependence. *Archives of General Psychiatry* 1996;53(3):217-24.
27. Krystal JH, Cramer JA, Krol W, et al. Efficacy of naltrexone for alcohol dependent veteran patients: VA Cooperative Study 425. *Alcoholism Clinical and Experimental Research* 2001;25(5 Supplement A):146A.
28. O'Malley S, Sinha R, Grilo C, et al. Naltrexone and cognitive behavioral coping skills therapy for the treatment of alcohol drinking and eating disorder features in alcohol-dependent women: A randomized controlled trial. *Alcoholism-Clinical and Experimental Research* 2007;31(4):625-34.
29. Gueorguieva R, Wu R, Pittman B, et al. New insights into the efficacy of naltrexone based on trajectory-based reanalyses of two negative clinical trials. *Biological Psychiatry* 2007;61(11):1290-5.
30. Kranzler H. Pharmacotherapy of alcoholism: Gaps in knowledge and opportunities for research. *Alcohol and Alcoholism* 2000;35(6):537-47.
31. McCaul ME, Petry NM. The role of psychosocial treatments in pharmacotherapy for alcoholism. *American Journal on Addictions* 2003;12:S41-S52.
32. Crits-Christoph P, Siqueland L, Blaine J, et al. Psychosocial treatments for cocaine dependence - National Institute on Drug Abuse Collaborative Cocaine Treatment Study. *Archives of General Psychiatry* 1999;56(6):493-502.
33. Carroll KM, Rounsaville BJ. A vision of the next generation of behavioral therapies research in the addictions. *Addiction* 2007;102(6):850-62.
34. Assanangkornchai S, Srisurapanont M. The treatment of alcohol dependence. *Current Opinion in Psychiatry* 2007;20(3):222-7.
35. Project Match Research G. Matching Alcoholism Treatments to Client Heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol* 1997;58(1):7-29.
36. Gordis E, Fuller R. Project MATCH. *Addiction* 1999;94(1):57-9.
37. Project Match Research G. Matching alcoholism treatments to client heterogeneity: Project MATCH three-year drinking outcomes. *Alcoholism Clinical and Experimental Research* 1998;22(6):1300-11.

38. Project Match Res G. Matching alcoholism treatments to client heterogeneity: Treatment main effects and matching effects on drinking during treatment. *Journal of Studies on Alcohol* 1998;59(6):631-9.
39. Orford J. Future research directions: a commentary on Project MATCH. *Addiction* 1999;94(1):62-6.
40. Allen JP, Mattson ME, Miller WR, et al. Summary of Project MATCH. *Addiction* 1999;94(1):31-4.
41. Project Match Research G. Project MATCH secondary a priori hypotheses. *Addiction* 1997;92(12):1671-806.
42. Cutler RB, Fishbain DA. Are alcoholism treatments effective? The project MATCH data. *Bmc Public Health* 2005;5.
43. Lindstrom L. Life is short, the Art long. *Addiction* 1999;94(1):45-7.
44. Karno M, Longabaugh R. What do we know? Process analysis and the search for a better understanding of project MATCH's anger-by-treatment matching effect. *Journal of Studies on Alcohol* 2004;65(4):501-12.
45. DiClemente CC. Mechanisms, determinants and processes of change in the modification of drinking behavior. *Alcoholism-Clinical and Experimental Research* 2007;31(10):13S-20S.
46. Longabaugh R, Wirtz PW, Zweben A, et al. Network support for drinking, Alcoholics Anonymous and long-term matching effects. *Addiction* 1998;93(9):1313-33.
47. Finney JW. Some treatment implications of Project MATCH. *Addiction* 1999;94(1):42-5.
48. Babor TF. Treatment for persons with substance use disorders: mediators, moderators, and the need for a new research approach. *International Journal of Methods in Psychiatric Research* 2008;17:S45-S9.
49. Allen J, Anton RF, Babor TF, et al. Matching alcoholism treatments to client heterogeneity: Project MATCH three-year drinking outcomes. *Alcoholism-Clinical and Experimental Research* 1998;22(6):1300-11.
50. Morgenstern J, Longabaugh R. Cognitive-behavioral treatment for alcohol dependence: a review of evidence for its hypothesized mechanisms of action. *Addiction* 2000;95(10):1475-90.
51. Longabaugh R, Donovan D, Karno M, et al. Active ingredients: How and why evidence-based alcohol behavioral treatment interventions work. *Alcoholism-Clinical and Experimental Research* 2005;29(2):235-47.
52. Longabaugh R, Zweben A, Locastro JS, et al. Origins, issues and options in the development of the combined behavioral intervention. *Journal of Studies on Alcohol* 2005:179-87.
53. Anton R, O'Malley S, Ciraulo D, et al. Combined pharmacotherapies and behavioral interventions for alcohol dependence - The COMBINE study: A randomized controlled trial. *Jama-Journal of the American Medical Association* 2006;295(17):2003-17.
54. Anton R, Randall C, Latham P, et al. Testing combined pharmacotherapies and behavioral interventions in alcohol dependence: Rationale and methods. *Alcoholism-Clinical and Experimental Research* 2003;27(7):1107-22.
55. Mason B. Acamprosate in the treatment of alcohol dependence. *Expert Opinion on Pharmacotherapy* 2005;6(12):2103-15.

56. Swift R, Pettinati H. Choosing pharmacotherapies for the COMBINE study - Process and procedures: An investigational approach to combination pharmacotherapy for the treatment of alcohol dependence. *Journal of Studies on Alcohol* 2005:141-7.
57. Ross S, Peselow E. Pharmacotherapy of Addictive Disorders. *Clinical Neuropharmacology* 2009;32(5):277-89.
58. Jung Y, Namkoong K. Pharmacotherapy for alcohol dependence: Anticraving medications for relapse prevention. *Yonsei Medical Journal* 2006;47(2):167-78.
59. Kiefer F, Jahn H, Tarnaske T, et al. Comparing and combining naltrexone and acamprosate in relapse prevention of alcoholism - A double-blind, placebo-controlled study. *Archives of General Psychiatry* 2003;60(1):92-9.
60. Weiss R, Kueppenbender K. Combining psychosocial treatment with pharmacotherapy for alcohol dependence. *Journal of Clinical Psychopharmacology* 2006;26:S37-S42.
61. Anton R, Moak D, Latham P, et al. Naltrexone combined with either cognitive behavioral or motivational enhancement therapy for alcohol dependence. *Journal of Clinical Psychopharmacology* 2005;25(4):349-57.
62. Gueorguieva R, Wu R, Donovan D, et al. Naltrexone and combined behavioral intervention effects on trajectories of drinking in the COMBINE study. *Drug and Alcohol Dependence* 2010;107(2-3):221-9.
63. Donovan D, Anton R, Miller W, et al. Combined pharmacotherapies and Behavioral interventions for alcohol dependence (The COMBINE study): Examination of posttreatment drinking outcomes. *Journal of Studies on Alcohol and Drugs* 2008;69(1):5-13.
64. Miller W, Moyers T, Arciniega L, et al. Training, supervision and quality monitoring of the COMBINE study behavioral interventions. *Journal of Studies on Alcohol* 2005:188-95.
65. Zhao X, Lynch JG, Jr., Chen Q. Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *Journal of Consumer Research* 2010;37(2):197-206.
66. Bergmark A. Interpretation of the outcome of the combine study: A response to the commentaries. *Addiction* 2008;103(5):710-.
67. Buehringer G, Pfeiffer-Gerschel T. Combine and match: The final blow for large-scale black box randomized controlled trials. *Addiction* 2008;103(5):708-10.
68. Kalant H. Drug treatment - Not as useless as it seems. *Addiction* 2008;103(5):707-8.
69. Mason B, Heyser C. The neurobiology, clinical efficacy and safety of acamprosate in the treatment of alcohol dependence. *Expert Opinion on Drug Safety* 2010;9(1):177-88.
70. Spanagel R, Zieglgansberger W. Anti-craving compounds for ethanol: New pharmacological tools to study addictive processes. *Trends in Pharmacological Sciences* 1997;18(2):54-9.
71. Cisler RA, Kivlahan DR, Donovan D, et al. Assessing nondrinking outcomes in combined pharmacotherapy and psychotherapy clinical trials for the treatment of alcohol dependence. *Journal of Studies on Alcohol* 2005:110-8.
72. Jaffe A, Rounsaville B, Chang G, et al. Naltrexone, relapse prevention, and supportive therapy with alcoholics: An analysis of patient treatment matching. *Journal of Consulting and Clinical Psychology* 1996;64(5):1044-53.
73. Monterosso JR, Flannery BA, Pettinati HM, et al. Predicting treatment response to naltrexone: The influence of craving and family history. *American Journal on Addictions* 2001;10(3):258-68.

74. Baron RM, Kenny DA. The Moderator Mediator Variable Distinction In Social Psychological-Research - Conceptual, Strategic, And Statistical Considerations. *Journal of Personality and Social Psychology* 1986;51(6):1173-82.
75. Nutt D. Alcohol and the brain - Pharmacological insights for psychiatrists. *British Journal of Psychiatry* 1999;175:114-9.
76. Mann K, Kiefer F, Spanagel R, et al. Acamprosate: Recent findings and future research directions. *Alcoholism-Clinical and Experimental Research* 2008;32(7):1105-10.
77. MacKinnon DP, Lockwood CM, Hoffman JM, et al. A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods* 2002;7(1):83-104.
78. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods* 2008;40(3):879-91.
79. Kaskutas LA, Bond J, Humphreys K. Social networks as mediators of the effect of Alcoholics Anonymous. *Addiction* 2002;97(7):891-900.
80. Kazdin AE, Nock MK. Delineating mechanisms of change in child and adolescent therapy: methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry and Allied Disciplines* 2003;44(8):1116-29.
81. Kelly JF, Abrantes AM, Brown SA. An 8-year follow-up of 12-step group involvement for treated adolescents. *Alcoholism-Clinical and Experimental Research* 2004;28(5):30A-A.
82. Williams J, Herman-Stahl M, Calvin SL, et al. Mediating mechanisms of a military web-based alcohol intervention. *Drug and Alcohol Dependence* 2009;100(3):248-57.
83. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface* 2009;2(4):457-68.
84. Petersen M, Sinisi S, van der Laan M. Estimation of direct causal effects. *Epidemiology* 2006;17(3):276-84.
85. van der Laan M, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data*. New York; 2011.
86. Mattson M, Litten R. Combining treatments for alcoholism: Why and how? *Journal of Studies on Alcohol* 2005:8-16.
87. Weiss RD, Locastro JS, Swift R, et al. Use of a "psychotherapy with no pills" treatment condition as part of a combined pharmacotherapy-psychotherapy research study of alcohol dependence. *Journal of Studies on Alcohol* 2005:43-9.
88. Kiefer F. Complete Suppression of Craving in Alcohol-Dependent Individuals Is It Possible? *Cns Drugs* 2009;23(5):361-7.
89. Anton R, Randall C, Latham P, et al. Testing combined pharmacotherapies and behavioral interventions for alcohol dependence (The COMBINE study): A pilot feasibility study. *Alcoholism-Clinical and Experimental Research* 2003;27(7):1123-31.
90. Ray L, Hutchison K. Effects of naltrexone on alcohol sensitivity and genetic moderators of medication response - A double-blind placebo-controlled study. *Archives of General Psychiatry* 2007;64(9):1069-77.
91. Monti PM, Rohsenow DJ, Hutchison KE, et al. Naltrexone's effect on cue-elicit craving among alcoholics in treatment. *Alcoholism-Clinical and Experimental Research* 1999;23(8):1386-94.

92. Kenna GA, McGeary JE, Swift RM. Pharmacotherapy, pharmacogenomics, and the future of alcohol dependence treatment, part 1. *American Journal of Health-System Pharmacy* 2004;61(21):2272-9.
93. Volpicelli J, Watson N, King A, Et Al. Effect Of Naltrexone On Alcohol High In Alcoholics. *American Journal of Psychiatry* 1995;152(4):613-5.
94. Soyka M, Helten C, Schmidt P. OCDS Craving Scores Predict 24-Month Outcome in Alcoholic Outpatients. *American Journal on Addictions* 2010;19(3):264-9.
95. Petrakis IL. Pharmacologic management of PTSD and comorbid alcohol dependence with naltrexone, disulfiram and antidepressants. *Alcoholism-Clinical and Experimental Research* 2006;30(6):184A-A.
96. Kranzler H, Armeli S, Feinn R, et al. Targeted naltrexone treatment moderates the relations between mood and drinking behavior among problem drinkers. *Journal of Consulting and Clinical Psychology* 2004;72(2):317-27.
97. Delongis A, Folkman S, Lazarus RS. The Impact Of Daily Stress On Health And Mood - Psychological And Social Resources As Mediators. *Journal of Personality and Social Psychology* 1988;54(3):486-95.
98. Torgerson DJ, Sibbald B. Understanding controlled trials - What is a patient preference trial? *British Medical Journal* 1998;316(7128):360-.
99. Meyer RE. Craving: what can be done to bring the insights of neuroscience, behavioral science and clinical science into synchrony. *Addiction* 2000;95(8):S219-S27.
100. Witkiewitz K, Bowen S, Donovan DM. Moderating Effects of a Craving Intervention on the Relation Between Negative Mood and Heavy Drinking Following Treatment for Alcohol Dependence. *Journal of Consulting and Clinical Psychology* 2011;79(1):54-63.
101. Bottlender M, Soyka M. Impact of craving on alcohol relapse during, and 12 months following, outpatient treatment. *Alcohol and Alcoholism* 2004;39(4):357-61.
102. Anton R, Kranzler H, Meyer R. Neurobehavioral Aspects Of The Pharmacotherapy Of Alcohol Dependence. *Clinical Neuroscience* 1995;3(3):145-54.
103. Capone C, Kahler C, Swift R, et al. Does Family History of Alcoholism Moderate Naltrexone's Effects on Alcohol Use? *Journal of Studies on Alcohol and Drugs* 2011;72(1):135-40.
104. Rohsenow DJ, Miranda R, Jr., McGeary JE, et al. Family history and antisocial traits moderate naltrexone's effects on heavy drinking in alcoholics. *Experimental and Clinical Psychopharmacology* 2007;15(3):272-81.
105. Krystal JH, Gueorguieva R, Cramer J, et al. Naltrexone is associated with reduced drinking by alcohol dependent patients receiving antidepressants for mood and anxiety symptoms: Results from VA cooperative study no. 425, "Naltrexone in the Treatment of Alcoholism". *Alcoholism-Clinical and Experimental Research* 2008;32(1):85-91.
106. McGeary JE, Monti PM, Rohsenow DJ, et al. Genetic moderators of naltrexone's effects on alcohol cue reactivity. *Alcoholism-Clinical and Experimental Research* 2006;30(8):1288-96.
107. Greenfield S, Pettinati H, O'Malley S, et al. Gender Differences in Alcohol Treatment: An Analysis of Outcome From the COMBINE Study. *Alcoholism-Clinical and Experimental Research* 2010;34(10):1803-12.
108. Kranzler HR, Tennen H, Armeli S, et al. Targeted Naltrexone for Problem Drinkers. *Journal of Clinical Psychopharmacology* 2009;29(4):350-7.

109. Anton R, Moak D, Waid L, et al. Naltrexone and cognitive behavioral therapy for the treatment of outpatient alcoholics: Results of a placebo-controlled trial. *American Journal of Psychiatry* 1999;156(11):1758-64.
110. Anton R. What is craving? Models and implications for treatment. *Alcohol Research & Health* 1999;23(3):165-73.
111. Petrakis I, Ralevski E, Nich C, et al. Naltrexone and disulfiram in patients with alcohol dependence and current depression. *Journal of Clinical Psychopharmacology* 2007;27(2):160-5.
112. Witkiewitz K, Bowen S. Depression, Craving, and Substance Use Following a Randomized Trial of Mindfulness-Based Relapse Prevention. *Journal of Consulting and Clinical Psychology* 2010;78(3):362-74.
113. Vieten C, Astin JA, Buscemi R, et al. Development of an Acceptance-Based Coping Intervention for Alcohol Dependence Relapse Prevention. *Substance Abuse* 2010;31(2):108-16.