A Descriptive Review and Meta-Regression of Demographic and Study Context Factors in US Clinical Trials of CBT Efficacy for Substance Use

Don-Pierre Burnett, Thomas A. Trikalinos, Chelsea Misquith, Molly Magill

Brown University School of Public Health, Providence RI

May, 2023

Thesis:

Submitted in partial fulfilment of the requirements for the

Degree of Master of Public Health in the Brown University School of Public Health

This research was supported by: NIAAA R01 AA029703 Awarded to Molly Magill

This thesis by Don-Pierre Burnett is accepted in its present form
by the Brown University School of Public Health as satisfying the
thesis requirements for the degree of Master of Public Health.

Date 4/28/2023

Molly Magill, PhD, Advisor

Date 4/28/2023

Thomas Trikalinos, MD, Reader

Date 4/28/2023

Annie Gjelsvik, PhD
Director, Master of Public Health Program

Approved by the Graduate Council

Date _____

_____
Thomas A. Lewis, Dean of the Graduate School

Preface and acknowledgements

I would first like to thank my advisor, Dr. Molly Magill for all of her support and mentorship throughout this project. I would like to thank Dr. Thomas Trikalinos and Chelsea Misquith for the integral roles they played on my thesis team and the wealth of expertise they contributed. Lastly, I would like to thank my friends and family, especially my mother Flolet, for their endless love and support.

Abstract

Background & Aims. Alcohol and other drug (AOD) use disorders are a significant public health concern. Cognitive-Behaviorally Based Interventions (CBIs) are frontline interventions for AOD treatment. This thesis study pursued three aims related to the efficacy of CBIs in terms of key population and study context variables. Firstly, we assessed the reporting of demographic information in these studies, to determine how demographic characteristics such as age, race, ethnicity, sex, gender, sexual orientation, and socioeconomic status are being reported and whether there is need for increased consideration of diversity in this research. Secondly, we examined CBI efficacy for different client sub-populations as defined by demographic characteristics. Finally, we examined CBI efficacy in different service contexts as defined by recruitment source (i.e., community samples, medical setting, specialty substance use or mental health setting, college setting, or criminal justice setting), treatment type (i.e., inpatient vs. outpatient), and treatment funding source (i.e., public vs. private). We also examined whether CBI efficacy differed by other study characteristic variables including drug type (i.e., alcohol, poly drug, cannabis, and other drugs [e.g. opiates and stimulants]), study inclusion of one or more co-occurring mental health conditions (i.e., yes or no), use of biological assay outcome (e.g., urine-screening) measures (i.e., yes or no), treatment length (i.e., scheduled number of sessions), CBT format (i.e. individual therapy, group, or mixed), and risk of bias (i.e. low risk vs. some concerns and high risk). **Methods.** Study characteristic data were collected and summarized using mean and standard deviation estimates as well as percentile values. Inverse-variance weighted standardized mean differences (effect sizes) were calculated for within- and between-group comparisons (i.e., CBI pre-test vs CBI post-test; CBI post-test vs control post-test, respectively) in each study and synthesized with random effects models. The demographic and study context variables used in the moderator analysis were determined during the data acquisition phase based on what was reported in the primary studies. Sensitivity analyses included assessing for study risk of bias using Cochrane guidelines, heterogeneity, influential studies, and publication bias. **Results.** The sample included K= 29 AOD clinical trials conducted in the United States (US). Of the racial/ethnic categories, the percentage of White participants was reported in most (26 studies [68% White on average]). Other racial/ethnic categories were less frequently reported and Native American, Asian, and Hawaiian Pacific Islander participants were underrepresented. The percentage of transgender participants was reported in three studies

and among those studies, 20% of participants were transgender. Four studies reported sexual orientation and among those studies, 59% of participants were heterosexual. The mean between-group effect size was small and moderately heterogenous (d = 0.158, 95% CI = 0.079, 0.238, p = 0.000, I^2 = 46.296) and the mean within-group effect large and very heterogenous (d_z = 1.147, 95% CI = 0.811, 1.482, p = 0.000, I^2 = 96.203). Drug type and biological assay outcome measures moderated between-group CBI efficacy and co-occurring disorders and study date moderated within-group CBI efficacy. **Conclusions.** Key demographic characteristics such as gender identity, sexual orientation, race, and ethnicity were underreported and/or were reported in non-inclusive ways. Non-alcohol studies had larger effects which may have been driven by large effects in two opioid studies. Studies with biological measures had larger effect sizes which may be explained by the superior precision of measurements or the overall robustness of these studies. Newer studies having smaller effects could potentially be a result of increased efforts to reduce publication bias. Co-occurring disorder studies having smaller effects may be due to the lower treatment adherence and higher drop-out often seen in these studies. Future research in the field of AOD treatment should seek to include more gender, sexual, racial, and ethnic minority participants, and report demographic information as inclusively and comprehensively as possible.

Introduction

Alcohol and other drug (AOD) use disorders are a serious public health concern. Alcohol use disorder (AUD) and substance use disorders (SUDs) are mental health conditions that lead to an inability to control one's use of alcohol, legal or illegal drugs, or prescription medications. AUD can cause short term health problems such as memory loss and black outs, as well as more serious long-term consequences such as cancer, liver cirrhosis, stomach problems, and increased risk of accidental death and suicide (Donroe & Edelman, 2022). In 2019 in the United States, 14.1 million adults over the age of 18 had AUD (National Survey on Drug Use and Health (NSDUH); SAMHSA, 2019). The 5th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) defines 11 symptoms of AUD and SUD that fall into the categories of impaired control, social problems, risky use, or physical dependence (American Psychiatric Association, 2013).[1] In addition, among US adults aged 18 – 25 in 2019, 14.1% (4.8 million people) had a SUD and among adults aged 26 and older, 6.7% (14.5 million people) had a SUD (NSDUH; SAMHSA, 2019). SUDs can have many negative consequences depending on the substance used, and they can co-occur with and contribute to the development of other mental health conditions (Hasin & Kilcoyne, 2012).

Cognitive Behavioral Therapy (CBT) is a form of intervention based on the premise that maladaptive cognitions, including general beliefs or schemas, contribute to the maintenance of emotional distress and behavioral problems (Hofmann et al., 2012). Marlatt and Gordon's (1985) work on Relapse Prevention laid the groundwork for the application of CBT for substance use treatment, and several important CBT manuals for AOD treatment subsequently followed (e.g., Carroll, 1998; Epstein & McCrady, 2009; Kadden et al., 1992; Monti et al., 1989). Previous reviews on CBT for substance use treatment have defined it as a time-bound, multi-session intervention targeting cognitive, affective, and environmental risk factors and teaching coping skills to promote positive outcomes (i.e., abstinence, reduction in use, or harm reduction, Magill et al., 2019; Magill et al., 2023). Cognitive behaviorally based interventions (CBIs) are widely

[1] The 11 symptoms are hazardous use, social and interpersonal problems related to use, neglecting major roles, withdrawal, tolerance, using larger amounts/longer, repeated attempts to control use/quit, much time spent using, physical or psychological problems related to use, activities given up to use, and craving (American Psychiatric Association, 2013). Displaying 2-3 symptoms indicates mild AUD/SUD, 4-5 indicate moderate AUD/SUD, and 6 or more indicates severe AUD/SUD, which is synonymous with addiction (American Psychiatric Association, 2013).

used treatment modalities for SUD (Ray et al., 2020). In a meta-analysis project that aggregated CBI efficacy data from studies dating as far back as 1990, Magill and colleagues (2019) found that CBIs were efficacious when compared to no or minimal treatment. CBIs were also efficacious when combined with pharmacotherapy, in comparison to standard medication management (Ray et al., 2020). CBIs did not however demonstrate efficacy over other evidence-based treatments such as Motivational Interviewing and Contingency Management (Magill et al., 2019; Ray et al., 2020).

Few studies have looked at how effect sizes from evidence-based interventions for AOD differ by sample demographic and other aspects of clinical trial context (i.e., contextual factors). In a review article on sex and gender differences in AUD recovery, Holzhauer and colleagues (2020) described that while there were mixed findings on the overall effect of sex and gender on AUD recovery, there is consistent evidence suggesting different mechanisms for behavior change between men and women in AUD treatment.[2] These authors found that overall, existing research on psychosocial treatments for AUD such as Motivational Enhancement Therapy and CBT, found no significant sex or gender differences (Holzhauer et al., 2020). The authors note that few studies directly test these sex and gender differences and that clinical trials for AUD often have low female enrollment (Holzhauer et al., 2020). They described the need for more research testing gender differences in treatment outcomes.

A meta-analysis conducted by Dutra and colleagues in 2008 examined the effects of psychosocial interventions for SUDs and found there was a moderate effect size. These authors examined demographic variables as moderators looking at the effects of age, sex, race (percent white), marital status (percent single/unmarried), employment status (percent employed part-time or full-time), and average years of substance use in relation to outcome variables such as treatment retention and effect size (Dutra et al., 2008). They found a large negative correlation between age and effect size, indicating that studies with younger samples were likely to have larger effect sizes (Dutra et al., 2008). They also found a negative correlation between average years of substance use and treatment dropout rate, indicating that patients with longer histories of

[2] Sex is a biological concept that describes differences within organisms related to reproductive, physiological, neuroendocrine, behavioral, and metabolic systems (Tannenbaum et al., 2019). Gender in comparison is psychologically, socially, and culturally constructed. Gender shapes attitudes, behaviors, and stereotypes and encompasses the interrelated concepts of gender norms, identity, and relations (Tannenbaum et al., 2019).

substance use were more likely to stay in treatment (Dutra et al., 2008). No other demographic variables were significant moderators in their study.

Windsor and Colleagues (2015) conducted a meta-analysis looking at differential efficacy of CBT by race and ethnicity, focusing on the differences in effect sizes between non-Hispanic White, Latino/Hispanic, and non-Hispanic Black participants. Their analysis included a sample of 16 studies on adult AOD published between 1997 and 2012. In their sample only 37.5% of studies specified the percentage Black participants, Latino/Hispanic participants, or other racial/ethnic groups. They did not find significant differences in the effect size of CBT versus comparison in studies with predominately non-Hispanic White participants (≥70%) versus studies with predominantly Black and/or Hispanic participants (Windsor et al., 2015). They did however find that the within-group pretest/post-test effects of CBT were significantly higher for studies with predominately non-Hispanic White participants compared to studies with predominantly Black and/or Hispanic participants (Hedges' g = 1.190, 0.861 respectively, Windsor et al., 2015). These authors stated that their results highlight the need for further research looking at the impact of CBIs on substance use outcomes among Black and Hispanic populations.

The present meta-analysis pursued three aims related to the efficacy of CBIs in relation to key population and study context variables. The first aim descriptively examined the reporting of demographic information in these studies, to determine how characteristics such as age, race, ethnicity, sex, gender, sexual orientation, and socioeconomic status are being reported and whether there is need for increased consideration of diversity in this research. The National Institutes of Health (NIH) policy on the inclusion of women and minorities in research clearly states that all clinical research must seek to include women and members of racial and ethnic minorities. In 2017, these guidelines were updated to require that researchers perform and report analyses by sex or gender, race, and/or ethnicity (National Institutes of Health, 2022). The second aim of this study examined if the efficacy of CBIs differs for different client sub-populations as defined by demographic characteristics (e.g., age, race, ethnicity, sex, gender, sexual orientation, and socioeconomic status). The third aim examined whether CBI efficacy differed by study context variables such as recruitment source (i.e., specialty substance use or mental health setting, community samples, college setting, criminal justice setting, or other setting), treatment type (i.e., outpatient vs. inpatient), and treatment funding source (i.e., private

vs. public). Aside from our primary aims, we also examined whether CBI efficacy differed by other study-level variables including drug type (i.e., alcohol, poly drug, cannabis, and other drugs [e.g. opiates and stimulants]), study inclusion of one or more co-occurring mental health conditions (i.e., yes or no), use of biological assay outcome (e.g., urine-screening) measures (i.e., yes or no), treatment length (i.e., scheduled number of sessions), CBT format (i.e. individual therapy, group, or mixed), and risk of bias (i.e. low risk vs. some concerns and high risk). Prior research has found that evidence-based treatment effect sizes are strongly related to the conditions they are compared to (e.g., no treatment versus another evidence-based treatment; Magill et al., 2019; Ray et al., 2020). We accounted for this by looking at between-group effect sizes across varying types of control conditions as well as within-group effect sizes for CBI from baseline to follow-up.

Methods

Primary Study Inclusion

Studies meeting the inclusion criteria for this meta-analysis were English-language, peer reviewed, randomized control trials (RCTS) published between 1980 and 2022. Studies were included if they were conducted in the United States, to allow for consistent comparisons of demographic and study-context factors. Studies included adults aged 18 and over, meeting criteria for AUD, SUD, or problematic use (DSM 3 through 5; Saunders et al., 1993). CBIs were the treatment of interest. These were either identified as CBT, Relapse Prevention (RP), or included strategies of CBT such as coping skills training, functional analysis, or high-risk situation avoidance. CBIs delivered in individual or group formats were included, but studies delivering CBT as an integrative therapy with other evidence-based interventions such as Mindfulness-Based Relapse Prevention were excluded. Interventions that were entirely technology-based were also excluded. All types of experimental control (i.e., minimal [waitlist, brief psychoeducation]; non-specific therapy [treatment as usual, drug counselling, supportive therapy]; specific therapy [Contingency Management, Motivational Interviewing]) were included as the type of control can directly impact the magnitude of the effect (Wampold, 2001), and within-group effects were calculated as well to provide a broad overview of CBI efficacy.

Literature Search

This project started with the dataset used in Magill et al., 2019. Magill and colleagues originally conducted a literature search for primary outcome reports of randomized controlled trials for AOD published from 1980 to 2018 (Details of this search can be found in Magill et al. 2019). A literature search was then conducted from July, 2018 through September, 2022 to update the dataset. Specifically, a title, abstract, and keyword search by treatment, outcome, and study terms was conducted in the PubMed database (full search strategy can be found in Appendix A). Other databases that were searched included Cochrane CENTRAL and the EBSCO databases (CINAHL, APA PsychINFO, and SocINDEX). Abstract screening was conducted by two raters through Covidence. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were followed, and study inclusion flow for this report is summarized in Figure 1. The final meta-analytic sample included K = 29 studies with 29 between-group and 23 within-group effect sizes.

[Figure 1]

Primary Study Characteristic Variables

This meta-analysis was interested in several study characteristic variables. All contrast condition types (i.e., minimal [waitlist, brief psychoeducation]; non-specific therapy [treatment as usual, drug counselling, supportive therapy]; specific therapy [Contingency Management, Motivational Interviewing]) were of interest in calculating between-group effect sizes. Within-group effect sizes were calculated using baseline and follow-up measures within CBI intervention groups. Early (1-6 months post-treatment) and late (7+ month post treatment) outcome data were collected, and late follow-up was used in analyses if early follow-up was not available.

Several study-level descriptors were defined a priori as potential meta regression covariates, and operationalized based on what was reported in the sample of primary studies. Study-level descriptors and potential meta-regression covariates were divided into the categories of demographic variables, service context variables, and other variables. Demographic variables included mean years of education, mean age, percentage of female participants, percentage or transgender participants, percentage of heterosexual participants, percentage of White participants, percentage of Black participants, percentage of Hispanic/Latino participants, percentage of Native American participants, percentage of Asian or Native Hawaiian/Pacific

Islander participants, percentage of participants identifying as more than one race, percentage of participants reporting some college education or more, and the percentage of participants with less than $20,000 annual income. Service context variables included study context (i.e., specialty substance use/mental health setting, community samples, college setting, criminal justice setting, or other setting), treatment funding source (i.e., private, or public), and treatment setting (i.e., outpatient, or inpatient). Other variables included drug type (i.e., alcohol, poly drug, cannabis, and other drugs [e.g. opiates and stimulants]), study inclusion of one or more co-occurring mental health conditions (i.e., yes or no), use of biological assay outcome (e.g., urine-screening) measures (i.e., yes or no), treatment length (i.e., scheduled number of sessions), CBT format (i.e. individual therapy, group, or mixed) and risk of bias (i.e. low risk, some concerns, or high risk). Data extraction occurred with two raters with an agreement rate for study codes of 98.4%. Data extraction guidelines were detailed in a study codebook which is available upon request.

Primary Study Outcome Variables

The standardized mean difference (Cohen's *d*) was used to measure between-group treatment efficacy outcomes in this analysis. Cohen's *d* for two groups of independent observations in a sample (d_s) is given by the following formula:

$$d_s = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1-1)SD_1^2+(n_2-1)SD_2^2}{n_1+n_2-2}}}$$

Figure 2. Cohen's d_s formula (Source: Lakens, 2013) $\bar{X}_i$ is the mean, SD_i the standard deviation, and n_i the sample size of the i-th group ($i = 1, 2$).

The standardized paired difference (Cohen's d_z) was used to measure within-group treatment efficacy outcomes. Cohen's d_z accounts for the correlation that exists between the dependent baseline and follow-up within-group measurements (Cohen, 1988; Lakens, 2013). Cohen's d_z is given by the following formula:

$$\text{Cohen's } d_z = \frac{M_{\text{diff}}}{\sqrt{\frac{\sum (X_{\text{diff}}-M_{\text{diff}})^2}{N-1}}}$$

Figure 3. Cohen's d_z formula (Source: Lakens, 2013) M_{diff} is the mean difference, and X_{diff} the pairwise differences for paired individuals in a sample of $N/2$ pairs (N people).

Effect sizes were weighted using inverse-variance weighting that minimizes the variance of an average by weighting each study's effect size in inverse proportion to its own variance (Deeks et al., 2022). This gives more-precise (usually, larger) studies that have smaller standard errors more weight than less precise (usually, smaller) studies with larger standard errors (Higgins & Green, 2011). Primary studies often reported data on multiple outcomes. The following decisional hierarchy was used to select data for use in effect size calculation: 1) biological assay measures, 2) frequency or quantity measures given as means and standard deviations, 3) sample proportions, and 4) other outcomes (i.e., the Addiction Severity Index; McLellan et al., 1980). Reverse scoring was used as needed to harmonize outcome measures with varying directionality (e.g., percentage of abstinent days versus percentage of use days) so that positive effect sizes corresponded to positive treatment outcomes.

Statistical Analysis

Primary study descriptive statistics were calculated using STATA version 17.0. Comprehensive Meta-analysis version 3.0 was used for all other analyses. Effect sizes were pooled under random effects assumptions. Whereas the fixed effects model assumes there is one true population effect size, the random effects model assumes there is systematic as well as random variability in the distribution of the population effect sizes (Higgins & Green, 2011). The benchmarks used to interpret effect size were: 0.2 "small", 0.5 "medium", and 0.8 "large" (Cohen, 1988). The significance of the Cochran's Q test was used to identify if statistically significant between-study heterogeneity existed and the I^2 statistic was used quantify the amount of heterogeneity (West et al., 2010). The magnitude of the I^2 statistic was interpreted as follows: 0 – 40% "may not be important", 30 – 60% "may show moderate heterogeneity", 50 – 90% "may show substantial heterogeneity", 75 – 100% "considerable heterogeneity" (Higgins & Green, 2011). Random effects meta-regressions were used to examine if demographic, service context, and other variables explained between-studies heterogeneity (Deeks et al., 2022). Variables were included in these analyses if they were reported in $K \geq 10$ studies to provide adequate statistical power (Deeks et al., 2022). Sensitivity analyses included assessing for study quality, influential studies, and publication bias. Study quality was rated for each study using the Cochrane Risk of Bias Tool (RoB 2; Higgins et al., 2011). Trimmed estimates were conducted where potentially influential studies are removed to see if the interpretation of the pooled effect will change in their absence (Higgins & Green, 2011). We also assessed for small-study effects

(systematic differences in effect size between less- and more-precise studies) using statistical tests and visually, by creating contour-enhanced funnel plots. Explanations for small study-effects include residual heterogeneity due to unidentified factors, chance, and publication and selective analysis reporting bias (Sterne et al., 2011).

Results

Primary Study Descriptive Characteristics

The sample included K= 29 of US-based AOD clinical trials. The study publication range was 1990 to 2022. The main substance use targets were alcohol (k = 10 [33%]), polydrug use (k = 9 [30%]), cannabis (k = 3 [10%]), and other drugs (including stimulants and opioids k = 8 [27%]). The mean age (SD) of participants in the sample was 37.67 (7.14). All 29 studies reported sex or gender with 33% of participants identifying as female. Sexual orientation was reported in 4 studies and among those studies, 59% of participants were heterosexual. The percentage of transgender participants was reported in 3 studies and among those studies, 20% of participants were transgender. Of the racial/ethnic categories, the percentage of White participants was reported most (k = 26 studies [68% White on average], followed by Black participants (k = 21 studies [37% Black on average]), and Hispanic/Latino participants (k = 15 studies [12% Hispanic/Latino on average]). Six studies reported the percentage of Native American, Asian or Hawaiian/Pacific Islander, and multiracial participants. Among those studies participants were 4% Native American, 1.5% Asian or Hawaiian/Pacific Islander, and 8% multiracial. Fourteen studies reported education as mean years and the average was 12.94 (0.82) years. Among the 21 studies that reported employment 47% of the patient samples were employed. Primary study descriptive statistics are detailed in Table 1. Study quality was assessed using Cochrane Risk of Bias tool; the results of which are summarized in Figure 4. Of the 29 studies in the sample 14 (48%) were rated as having a high risk of bias, 3 (10%) were rated as having some concerns, and 11 (38%) were rated as having a low risk of bias.

[Table 1]

CBT Effect by Comparison Type

Between-group effects. Between- and within-group effect sizes at early follow-up are detailed in Table 2. Twenty-nine US-based clinical trials contributed between-group effect sizes, and the pooled effect d = 0.158 (95% CI = 0.079, 0.238, $p < 0.0005$, $I^2 = 46\%$). This is a small

effect with moderate potential heterogeneity. Trimmed estimates, where each primary study is removed one at a time to test for studies with a large influence on the overall effect size, showed no influential studies. The funnel plot was however asymmetrical, and the rank order correlation was positive and significant which indicated that there was potential publication bias (Figure 5; $\tau = .29$, $p = 0.027$, Figure 9).

Within-group effects. Twenty-three US-based studies contributed within-group effect sizes, which describe the effects of CBT within the intervention group from baseline to early follow-up. The pooled effect $d_z = 1.147$ (95% CI = 0.811, 1.482, $p < 0.0005$, $I^2 = 96\%$). This is a large effect size with considerable heterogeneity based on the large I^2. Trimmed estimates showed that no one study had a disproportionate impact on the overall effect size. The funnel plot showed some asymmetry, but the rank order correlation was not significant so overall there was little evidence supporting potential publication bias (Figure 6; $\tau = 0.16$, $p = 0.279$, Figure 10).

[Table 2]

Meta-regression Analyses

Meta-regression analyses were used to assess if the pooled between- and within-group effect sizes for CBIs were moderated by demographic, service context, and other variables. Results from these analyses are presented in Table 3.

Between-group analysis. In the between-group analysis drug type was a significant effect size moderator. Compared to alcohol-focused treatment studies, non-alcohol studies had larger effect sizes ($\beta = 0.1829$, $p = 0.0127$). Biological assay outcome measures were a significant moderator, and studies without them had smaller effect sizes ($\beta = -0.2356$, $p = 0.0216$). No demographic or service context variables were significant moderators in this model.

Within-group analysis. In the within-group model study date was a significant moderator, and newer studies had smaller effect sizes ($\beta = -0.0492$, $p = 0.0013$). Co-occurring disorder was also a significant moderator, and co-occurring disorder studies had smaller effect sizes ($\beta = -0.7276$, $p = 0.0342$). No demographic or service context variables were significant in this model.

Several variables of interest were not included in these analyses because they were reported in too few studies (K < 10). The variables that were not analyzed included the percentage of transgender participants, percentage of heterosexual participants, percentage of Native American participants, percentage of Asian or Native Hawaiian/Pacific Islander participants, percentage of participants with some college education or more, and the percentage of participants with less than $20,000 annual income.

[Table 3]

Discussion

The aims of this meta-analysis were to (a) examine the reporting of demographic information in primary studies; (b) examine if CBI efficacy and effectiveness differed for different client-subpopulations as defined by demographic characteristics; and (c) examine if CBI efficacy and effectiveness differed by service context as defined by recruitment source, treatment setting, and treatment funding source. We also examined if the efficacy of CBI differed by other study factors including drug type, co-occurring diagnosis, biological assay measures, treatment length, study date, and CBI format. We found that key demographic variables such as gender identity, sexual orientation, race, and ethnicity were underreported and/or reported in non-inclusive ways.

Most studies did not disambiguate between biological sex and gender which is multifaceted and psychosocially constructed. Gender was reported solely as the number/percentage of male/female participants in 26 of the 29 studies reviewed. Only three studies, all of which were published in 2019 or later, reported the percentage of transgender participants. Similarly, sexual orientation was only reported in 4 studies. Moore and colleagues (2019) reported the percentage of male, female, and transgender participants, but did not report sexual orientation. In their intervention for depression, anxiety, and alcohol use among gender diverse and sexual minority women, Pachankis and colleagues (2021) reported the percentage of cisgender and transgender women. They also reported sexual orientation categorically as lesbian, queer, and other (bisexual, pansexual, asexual, uncertain). In their 2021 study on substance use and medication adherence among Black/African American individuals living with HIV/AIDS, Magidson and colleagues reported the percentages of transgender and sexual minority participants. Finally, in their 2022 study on CBT for heavy drinking among individuals with HIV,

Dévieux and colleagues reported gender as the percentage of male and female participants, and reported sexual orientation categorically as percent homosexual, heterosexual, bisexual, none of the above, and refused to answer. The inconsistency of this reporting made it difficult to operationalize gender and sexual orientation in more inclusive ways. The reporting of gender and sexual orientation in the Pachankis et al. (2021) paper was the most comprehensive. Overall, the underreporting and underrepresentation of gender and sexual minority participants is concerning, considering the increased risk of AUD and SUD faced by these populations. According to the 2020 National Survey on Drug Use and Health (NSDUH) 21.8% of sexual minority adults had an alcohol use disorder compared to 11% in the general population (NSDUH, 2020). Transgender individuals face rates of SUD as high as 28%, which have been associated with experiences of abuse, stigma, and discrimination (Oberheim et al., 2017).

Disparities were also seen in the reporting and representation of several racial/ethnic groups. The percentage of White participants was most frequently reported. Only 6 studies reported the percentage of Native American participants. In those 6 studies, participants were 4.31% (range: 0.9% to 7%) Native American, meaning the overall representation of Native Americans was very low. AUD and SUD are serious public health concerns in Native American populations, which makes this underrepresentation alarming despite calls for increased representation in clinical research (Mainous et al., 2023). Asian and Native Hawaiian/Pacific Islander were also underrepresented racial/ethnic groups. Only 6 studies reported these categories, and their overall representation was 1.51%. The US population is 6.1% Asian and 0.3% Native Hawaiian/Pacific Islander, so these groups were very underrepresented. It was also unfortunate that these categories are often grouped together in primary studies.

Our analysis did not find differences in CBI efficacy based on demographic or service context variables. Lack of reporting and representation in primary studies resulted in several variables of interest being excluded from the moderator analysis. Lack of statistical power could also play a role in some of the null findings observed here, suggesting that researchers should aim to recruit more diverse samples and report demographics more inclusively. NIH guidelines require that researchers seek to recruit women and racial/ethnic minorities and perform and report analyses by sex or gender, race, and/or ethnicity (National Institutes of Health, 2022). We believe researchers should go above and beyond these guidelines and strive to recruit more gender, sexual, racial, and ethnic minority participants and report this demographic data in the

most inclusive ways possible. We did find several variables that were significant moderators of CBT efficacy.

The between-group pooled effect size of CBT was d = 0.158 (95% CI = 0.079, 0.238, p = 0.000, I^2 = 46.296). In the analysis between CBIs and control conditions, drug type was a significant moderator, and non-alcohol studies had larger effect sizes. Ray et al., (2020) also found that primary drug type was a significant moderator of between-study heterogeneity, however they found that alcohol studies had larger effects. These results could be driven by moderate to large effect sizes in a select few opioid studies, specifically McAuliffe et al., (1990) and Barry et al., (2019). Biological assay outcome measures were a significant moderator, and studies with biological measures had larger effect sizes. This could have to do with these studies having more resources or may be due to the superior precision of measurements.

The within-group pooled effect size was d_z = 1.147 (95% CI = 0.811, 1.482, p = 0.000, I^2 = 96.203). In the within-group analysis study date and dual diagnosis were significant moderators. Newer studies having smaller effects could potentially be a result of increased efforts to reduce publication bias in research. Co-occurring disorder studies having smaller effects is an interesting finding. Psychiatric disorder and AOD are highly comorbid, and studies have consistently shown that treating co-occurring disorders simultaneously is more effective than treating them individually (Kelly & Daley, 2013). Co-occurring psychiatric disorders and AOD have however been associated with lower treatment adherence and higher drop-out, which could contribute to our findings (Demarce et al., 2008; Kelly & Daley, 2013). The added complexity of treating co-occurring AOD and physical or mental health conditions simultaneously could also be a factor.

Limitations

There were several limitations worth noting in this analysis. Of our three stated aims, two relied on conducting moderator analyses, but due to limitations in primary study reporting these analyses were potentially underpowered. There was also significant heterogeneity in the within-group pooled effect, and based on our regressions it was not substantially explained by the moderators we considered. Based on the R^2 estimates in the within-group analysis, the only moderator that explained a significant amount of variance was study date in relation to the within-group CBI effect size (R^2 = 0.18). Several unmeasured factors could account for this

heterogeneity including clinical differences between studies, methodological differences, inconsistent measurement of moderator variables, and publication bias (Fletcher, 2007). The exclusion of technologically-based CBIs is also a potential limitation of this study. The decision to exclude these interventions was made as we wanted homogeneity in how CBI was defined and delivered. There is however an increasing shift towards web and app-based interventions, and they do have potential for reducing barriers and increasing treatment access (Hadjistavropoulos et al., 2020). In view of this, it is important that technologically based CBIs for AOD be considered, and it is a limitation that we were not able to look at them in this review. The focus on consumption measures was also a limitation. Optimal outcomes in clinical trials for AUD/SUD are not clearcut. Reductions in negative consequences from substance use and improvements in daily functioning can be just as important as reducing consumption (Kiluk et al., 2019). These were however outside the scope of this review. The presence of publication bias in the between-group pooled effect is another limitation of this study. Lastly, we only examined early follow-up (1-6 months post treatment). It would have been ideal to be able to conduct these analyses with late follow-up data as well.

Conclusion

This analysis found that key demographic characteristics such as gender identity, sexual orientation, race, and ethnicity were underreported in US-based clinical trials of CBI interventions and/or were reported in non-inclusive ways with too few categories. The representation of specific groups such as gender/sexually diverse, Native American, Asian, and NH/PI individuals was noticeably low, as this is particularly concerning as several of these populations are at elevated risk for AOD. Our moderator analysis did not find significant differences in treatment effects due to demographic and study contextual characteristics and sub-groups of interest were excluded from the analysis due to insufficient reporting and representation in primary studies. Drug type and biological assay outcome measures did moderate between-group CBI efficacy. Co-occurring disorders and study date were moderators of within-group CBI efficacy. This was also the first study to look at the within-group efficacy of CBIs in an entirely US sample. Future studies on CBIs for AUD/SUD should strive to include more diverse samples. Demographic characteristics such as race, ethnicity, gender identity, and sexual orientation should also be reported more consistently and inclusively.

References

References marked with an asterisk (*) are included in the meta-analysis.

American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders (3rd ed., test rev.)*. Washington, DC, USA: American Psychiatric Association.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC, USA: American Psychiatric Association.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th Ed.). Washington, DC, USA: American Psychiatric Association.

Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic

*Barry, D. T., Beitel, M., Cutter, C. J., Fiellin, D. A., Kerns, R. D., Moore, B. A., Oberleitner, L., Madden, L. M., Liong, C., Ginn, J., & Schottenfeld, R. S. (2019). An evaluation of the feasibility, acceptability, and preliminary efficacy of cognitive-behavioral therapy for opioid use disorder and chronic pain. *Drug and Alcohol Dependence, 194*, 460–467. https://doi.org/10.1016/j.drugalcdep.2018.10.015

*Bowen, S., Witkiewitz, K., Clifasefi, S. L., Grow, J., Chawla, N., Hsu, S. H., Carroll, H. A., Harrop, E., Collins, S. E., Lustyk, M. K., & Larimer, M.E. (2014). Relative efficacy of mindfulness-based relapse prevention, standard relapse prevention, and treatment as usual for substance use disorders: A randomized clinical trial. *JAMA Psychiatry, 71*(5), 547-556. https://dx.doi.org/10.1001/jamapsychiatry.2013.4546

*Brown, T. G., Seraganian, P., Tremblay, J., & Annis, H. (2002). Matching substance abuse aftercare treatments to client characteristics. *Addictive Behaviors, 27*(4), 585-604. https://dx.doi.org/10.1016/S0306-4603(01)00195-2

*Budney, A. J., Moore, B. A., Rocha, H. L., & Higgins, S. T. (2006). Clinical trial of abstinence-based vouchers and cognitive-behavioral therapy for cannabis dependence. *Journal of Consulting and Clinical Psychology, 74*(2), 307-316. https://dx.doi.org/10.1037/0022-006X.74.2.307

Carroll, K. M. (1998). A cognitive behavioral approach: Treating cocaine addiction. NIH Publication Number: 984308. Rockville, MD: National Institute on Drug Abuse.

*Carroll, K. M., Rounsaville, B. J., & Gawin, F. H. (1991). A comparative trial of psychotherapies for ambulatory cocaine abusers: relapse prevention and interpersonal

psychotherapy. *The American Journal of Drug and Alcohol Abuse, 17*(3), 229-247. https://dx.doi.org/10.3109/00952999109027549

Deeks J.J., Higgins J.P.T., Altman D.G. (2002). Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins J.P.T., Thomas J., Chandler J., Cumpston M., Li T., Page M.J., Welch V.A. *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane, 2022. Retrieved June 8, 2022, from www.training.cochrane.org/handbook

DeMarce, J. M., Lash, S. J., Stephens, R. S., Grambow, S. C., & Burden, J. L. (2008). Promoting continuing care adherence among substance abusers with co-occurring psychiatric disorders following residential treatment. Addictive Behaviors, 33(9), 1104–1112. https://doi.org/10.1016/j.addbeh.2008.02.008

*Dévieux, J. G., Rosenberg, R., Jean-Gilles, M., Villalba, K., Attonito, J., Lerner, B., Saxena, A., & Stein, J. (2022). Effectiveness of a cognitive behavioral randomized controlled trial for people living with HIV who are heavy drinkers: The Holistic Health Recovery Program (HHRP) trial in Miami. *Journal of Clinical Psychology in Medical Settings*, *29*(3), 498–508. https://doi.org/10.1007/s10880-022-09878-5

Donroe, J. H., & Edelman, E. J. (2022). Alcohol use. *Annals of Internal Medicine*, *175*(10). https://doi.org/10.7326/aitc202210180

Dutra, L., Stathopoulou, G., Basden, S. L., Leyro, T. M., Powers, M. B., & Otto, M. W. (2008). A meta-analytic review of psychosocial interventions for substance use disorders. *The American journal of psychiatry*, *165*(2), 179–187. https://doi.org/10.1176/appi.ajp.2007.06111851

*Easton, C. J., Crane, C. A., & Mandel, D. (2017). A randomized controlled trial assessing the efficacy of cognitive behavioral therapy for substance-dependent domestic violence offenders: An integrated substance abuse-domestic violence treatment approach (SADV). *Journal of Marital and Family Therapy*, *44*(3), 483–498. https://doi.org/10.1111/jmft.12260

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629-634. doi:10.1136/bmj.315.7109.629

Epstein, E. E., & McCrady, B. S. (2009). *A cognitive-behavioral treatment program for overcoming alcohol problems: Therapist guide*. Oxford, UK: Oxford University Press.

Fletcher J. (2007). What is heterogeneity and is it important?. *BMJ (Clinical research ed.)*, *334*(7584), 94–96. https://doi.org/10.1136/bmj.39057.406644.68

Hadjistavropoulos, H. D., Mehta, S., Wilhelms, A., Keough, M. T., & Sundström, C. (2019). A systematic review of internet-delivered cognitive behavior therapy for alcohol misuse: Study characteristics, program content and outcomes. *Cognitive Behaviour Therapy*, *49*(4), 327–346. https://doi.org/10.1080/16506073.2019.1663258

Hasin, D., & Kilcoyne, B. (2012). Comorbidity of psychiatric and substance use disorders in the United States: current issues and findings from the NESARC. *Current opinion in psychiatry*, *25*(3), 165–171. https://doi.org/10.1097/YCO.0b013e3283523dcc

Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, … & Sterne JA (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 343, d5928 10.1136/bmj.d5928 [PubMed: 22008217]

Hofmann, S. G., Asnaani, A., Vonk, I. J., Sawyer, A. T., & Fang, A. (2012). The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. *Cognitive therapy and research*, *36*(5), 427–440. https://doi.org/10.1007/s10608-012-9476-1

Holzhauer, C. G., Cucciare, M., & Epstein, E. E. (2020). Sex and Gender Effects in Recovery From Alcohol Use Disorder. *Alcohol research : current reviews*, *40*(3), 03. https://doi.org/10.35946/arcr.v40.3.03

*Ilgen, M. A., Coughlin, L. N., Bohnert, A. S., Chermack, S., Price, A., Kim, H. M., Jannausch, M., & Blow, F. C. (2020). Efficacy of a psychosocial pain management intervention for men and women with substance use disorders and chronic pain. *JAMA Psychiatry*, *77*(12), 1225. https://doi.org/10.1001/jamapsychiatry.2020.2369

Kadden, R. M., Carroll, K., Donovan, D., Cooney, N., Monti, P., Abrams, D., Litt, M. & Hester, R. (Eds.) (1992). Cognitive-Behavioral Coping Skills Therapy Manual: A Clinical Research Guide for Therapists Treating Individuals with Alcohol Abuse and Dependence. Volume 4, Project MATCH Monograph Series. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism. (DHHS Publication No. (ADM)92-1895).

*Kadden, R. M., Litt, M. D., Cooney, N. L., Kabela, E., & Getter, H. (2001). Prospective matching of alcoholic clients to cognitive-behavioral or interactional group

therapy. *Journal of Studies on Alcohol*, *62*(3), 359-369. https://dx.doi.org/10.15288/jsa.2001.62.359

Kelly, T. M., & Daley, D. C. (2013). Integrated treatment of substance use and psychiatric disorders. *Social Work in Public Health*, *28*(3-4), 388–406. https://doi.org/10.1080/19371918.2013.774673

Kiluk, B. D., Fitzmaurice, G. M., Strain, E. C., & Weiss, R. D. (2019). What Defines a Clinically Meaningful Outcome in the Treatment of Substance Use Disorders: Reductions in Direct Consequences of Drug Use or Improvement in Overall Functioning?. *Addiction*, *114*(1), 9-15. https://dx.doi.org/10.1111/add.14289

*Kivlahan, D. R., Marlatt, G. A., Fromme, K., Coppel, D. B., & Williams, E. (1990). Secondary prevention with college drinkers: Evaluation of an alcohol skills training program. *Journal of Consulting and Clinical Psychology*, *58*(6), 805-810. https://dx.doi.org/10.1037/0022-006X.58.6.805

*Kushner, M. G., Maurer, E. W., Thuras, P., Donahue, C., Frye, B., Menary, K. R., Hobbs, J., Haeny, A. M., & Van Demark, J. (2013). Hybrid cognitive behavioral therapy versus relaxation training for co-occurring anxiety and alcohol disorder: a randomized clinical trial. *Journal of consulting and clinical psychology*, *81*(3), 429–442. https://doi.org/10.1037/a0031301

Lakens D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology*, 4, 863. https://doi.org/10.3389/fpsyg.2013.00863

*Litt, M. D., Kadden, R. M., Cooney, N. L., & Kabela, E. (2003). Coping skills and treatment outcomes in cognitive-behavioral and interactional group therapy for alcoholism. *Journal of Consulting and Clinical Psychology*, *71*(1), 118–128. https://doi.org/10.1037/0022-006x.71.1.118

*Litt, M. D., Kadden, R. M., Tennen, H., & Kabela-Cormier, E. (2016). Network Support II: Randomized controlled trial of Network Support treatment and cognitive behavioral therapy for alcohol use disorder. *Drug and Alcohol Dependence*, *165*, 203-212. https://dx.doi.org/10.1016/j.drugalcdep.2016.06.010

*Magidson, J. F., Belus, J. M., Seitz-Brown, C. J., Tralka, H., Safren, S. A., & Daughters, S. B. (2021). Act healthy: A randomized clinical trial evaluating a behavioral activation

intervention to address substance use and medication adherence among low-income, Black/African American individuals living with HIV/AIDS. *AIDS and Behavior*, *26*(1), 102–115. https://doi.org/10.1007/s10461-021-03354-1

Magill, M., Ray, L., Kiluk, B., Hoadley, A., Bernstein, M., Tonigan, J. S., & Carroll, K. (2019). A meta-analysis of cognitive-behavioral therapy for alcohol or other drug use disorders: Treatment efficacy by contrast condition. *Journal of Consulting and Clinical Psychology*, *87*(12), 1093–1105. https://doi.org/10.1037/ccp0000447

Magill, M., Kiluk, B. D., & Ray, L. A. (2023). Efficacy of Cognitive Behavioral Therapy for Alcohol and Other Drug Use Disorders: Is a One-Size-Fits-All Approach Appropriate?. *Substance Abuse and Rehabilitation*, 1-11.

Magill, M., & Ray, L. A. (2009). Cognitive-behavioral treatment with adult alcohol and illicit drug users: a meta-analysis of randomized controlled trials. *Journal of studies on alcohol and drugs*, *70*(4), 516–527. https://doi.org/10.15288/jsad.2009.70.516

Mainous, A. G., Kelliher, A., & Warne, D. (2023). Recruiting indigenous patients into clinical trials: A circle of trust. *The Annals of Family Medicine*, *21*(1), 54–56. https://doi.org/10.1370/afm.2901

Marlatt, G. A., & Gordon, J. R. (Eds.). (1985). *Relapse prevention: Maintenance strategies in the treatment of addictive behaviors.* New York: Guilford Press.

*Maude-Griffin, P.M., Hohenstein, J.M., Humfleet, G.L., Reilly, P.M., Tusel, D.J., & Hall, S.M. (1998). Superior efficacy of cognitive-behavioral therapy for urban crack cocaine abusers: Main and matching effects. *Journal of Consulting and Clinical Psychology, 66*(5), 832-837. https://dx.doi.org/10.1037/0022-006X.66.5.832

*McAuliffe, W. E. (1990). A randomized controlled trial of recovery training and self-help for opioid addicts in New England and Hong Kong. *Journal of Psychoactive Drugs*, *22*(2), 197-209. https://dx.doi.org/10.1080/02791072.1990.10472544

*McGovern, M. P., Lambert-Harris, C., Alterman, A. I., Xie, H., & Meier, A. (2011). A randomized controlled trial comparing integrated cognitive behavioral therapy versus individual addiction counseling for co-occurring substance use and posttraumatic stress disorders. *Journal of Dual Diagnosis*, *7*(4), 207–227. https://doi.org/10.1080/15504263.2011.620425

*McGovern, M. P., Lambert-Harris, C., Xie, H., Meier, A., McLeman, B., & Saunders, E. (2015). A randomized controlled trial of treatments for co-occurring substance use disorders and post-traumatic stress disorder. *Addiction*, *110*(7), 1194–1204. https://doi.org/10.1111/add.12943

*McKay, J. R., Alterman, A. I., Cacciola, J. S., Rutherford, M. J., O'Brien, C. P., & Koppenhaver, J. (1997). Group counseling versus individualized relapse prevention aftercare following intensive outpatient treatment for cocaine dependence: Initial results. *Journal of Consulting and Clinical Psychology, 65*(5), 778-788. https://dx.doi.org/10.1037/0022-006X.65.5.778

*McKay, J. R., Lynch, K. G., Coviello, D., Morrison, R., Cary, M. S., Skalina, L., & Plebani, J. (2010). Randomized trial of continuing care enhancements for cocaine-dependent patients following initial engagement. *Journal of Consulting and Clinical Psychology, 78*(1), 111-120. https://dx.doi.org/10.1037/a0018139

*McKay, J. R., Lynch, K. G., Shepard, D. S., Ratichek, S., Morrison, R., Koppenhaver, J., & Pettinati, H. M. (2004). The effectiveness of telephone-based continuing care in the clinical management of alcohol and cocaine use disorders: 12-month outcomes. *Journal of Consulting and Clinical Psychology, 72*(6), 967-979. https://dx.doi.org/10.1037/0022-006X.72.6.967

McLellan AT, Luborsky L, Woody GE, O'Brien CP. An improved diagnostic evaluation instrument for substance abuse patients. The Addiction Severity Index. *J Nerv Ment Dis*. 1980;168(1):26-33. doi:10.1097/00005053-198001000-00006

Monti, P. M., Abrams, D. B., Kadden, R. M., & Cooney, N. L. (1989). Treating Alcohol Dependence: A Coping Skills Training Guide. New York: Guilford Press.

*Monti, P. M., Rohsenow, D. J., Michalec, E., Martin, R. A., & Abrams, D.B. (1997). Brief coping skills treatment for cocaine abuse: Substance use outcomes at three months. Addiction, *92*(12), 1717-1728. https://dx.doi.org/10.1111/j.1360-0443.1997.tb02892.x

*Moore, A. A., Lake, J. E., Glasner, S., Karlamangla, A., Kuerbis, A., Preciado, D., Jenkins, J., Dominguez, B. X., Candelario, J., Liao, D. H., Tang, L., & Reid, M. C. (2019). Establishing the feasibility, acceptability and preliminary efficacy of a multi-component behavioral intervention to reduce pain and substance use and improve physical

performance in older persons living with HIV. *Journal of Substance Abuse Treatment*, *100*, 29–38. https://doi.org/10.1016/j.jsat.2019.02.003

Nathan, P. E., Horvath, T., Zweben, J., & Wallace, J. (n.d.). *Understanding Alcohol Use Disorders and their treatment*. American Psychological Association. Retrieved May 15, 2022, from https://www.apa.org/topics/substance-use-abuse-addiction/alcohol-disorders

National Institutes of Health. (2021, April). *Understanding alcohol use disorder*. National Institute on Alcohol Abuse and Alcoholism. Retrieved May 7, 2022, from https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/understanding-alcohol-use-disorder

National Institutes of Health. (2022, October 11). *Inclusion of women and minorities as participants in research involving human subjects*. National Institutes of Health. Retrieved April 10, 2023, from https://grants.nih.gov/policy/inclusion/women-and-minorities.htm

National Institute of Mental Health. (n.d.). *Substance use and co-occurring mental disorders*. National Institute of Mental Health. Retrieved May 18, 2022, from https://www.nimh.nih.gov/health/topics/substance-use-and-mental-health#:~:text=A%20substance%20use%20disorder%20(SUD,most%20severe%20form%20of%20SUDs.

Oberheim, S. T., DePue, M. K., & Hagedorn, W. B. (2017). Substance use disorders (SUDS) in transgender communities: The need for trans-competent sud counselors and facilities. *Journal of Addictions & Offender Counseling*, *38*(1), 33–47. https://doi.org/10.1002/jaoc.12027

*Osilla, K. C., Paddock, S. M., McCullough, C. M., Jonsson, L., & Watkins, K. E. (2019). Randomized clinical trial examining cognitive behavioral therapy for individuals with a first-time DUI offense. *Alcoholism: Clinical and Experimental Research*, *43*(10), 2222–2231. https://doi.org/10.1111/acer.14161

*Pachankis, J. E., McConocha, E. M., Clark, K. A., Wang, K., Behari, K., Fetzner, B. K., Brisbin, C. D., Scheer, J. R., & Lehavot, K. (2020). A transdiagnostic minority stress intervention for gender diverse sexual minority women's depression, anxiety, and unhealthy alcohol use: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *88*(7), 613–630. https://doi.org/10.1037/ccp0000508

Pitpitan, E. V., Semple, S. J., Zians, J., Strathdee, S. A., & Patterson, T. L. (2018). Mood, meth, condom use, and gender: Latent growth curve modeling results from a randomized trial. *AIDS and Behavior*, *22*(9), 2815–2829. https://doi.org/10.1007/s10461-018-2147-9

*Project MATCH Research Group. (1997) Matching Alcoholism Treatments to Client Heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol, 58*(1), 7-29. https://dx.doi.org/10.15288/jsa.1997.58.7

Ray, L. A., Meredith, L. R., Kiluk, B. D., Walthers, J., Carroll, K. M., & Magill, M. (2020). Combined Pharmacotherapy and Cognitive Behavioral Therapy for Adults With Alcohol or Substance Use Disorders: A Systematic Review and Meta-analysis. *JAMA network open*, *3*(6), e208279. https://doi.org/10.1001/jamanetworkopen.2020.8279

Saunders JB, Aasland OG, Babor TF, De la Fuente JR, & Grant M (1993). Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction*, 88(6), 791–804. 10.1111/j. 1360-0443.1993.tb02093.x [PubMed: 8329970]

Shorten, A., & Shorten, B. (2012). What is meta-analysis? *Evidence Based Nursing*, *16*(1), 3–4. https://doi.org/10.1136/eb-2012-101118

*Stephens, R. S., Roffman, R. A., & Curtin, L. (2000). Comparison of extended versus brief treatments for marijuana use. *Journal of Consulting and Clinical Psychology*, *68*(5), 898-908. https://dx.doi.org/10.1037/0022-006X.68.5.898

*Stephens, R. S., Roffman, R. A., & Simpson, E. E. (1994). Treating adult marijuana dependence: A test of the relapse prevention model. *Journal of Consulting and Clinical Psychology*, *62*(1), 92-99. https://dx.doi.org/10.1037/0022-006X.62.1.92

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ (Clinical research ed.)*, *343*, d4002. https://doi.org/10.1136/bmj.d4002

Substance Abuse and Mental Health Services Administration. (n.d.). *Key substance use and mental health indicators in the United States:results from the 2019 National Survey on Drug Use and health*. Key Substance Use and Mental Health Indicators in the United

States: Results from the 2019 National Survey on Drug Use and Health. Retrieved May 20, 2022, from https://www.samhsa.gov/data/sites/default/files/reports/rpt29393/2019NSDUHFFRPDFWHTML/2019NSDUHFFR090120.htm#sud

Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health (NSDUH). (2019). *Table 5.4A—Alcohol Use Disorder in Past Year Among Persons Aged 12 or Older, by Age Group and Demographic Characteristics: Numbers in Thousands, 2018 and 2019.* Rockville, MD: Substance abuse and Mental Health Services Administration. Retrieved May 20, 2022, from https://www.samhsa.gov/data/sites/default/files/reports/rpt29394/NSDUHDetailedTabs2019/NSDUHDetTabsSect5pe2019.htm?s=5.4&#tab5-4a

Tannenbaum, C., Ellis, R. P., Eyssel, F., Zou, J., & Schiebinger, L. (2019). Sex and gender analysis improves science and engineering. *Nature, 575*(7781), 137–146. https://doi.org/10.1038/s41586-019-1657-6

*Thornton, C., Gottheil, E., Patkar, A., & Weinstein, S. (2003). Coping styles and response to high versus low-structure individual counseling for substance abuse. *American Journal on Addictions, 12*(1), 29-42. https://dx.doi.org/10.1111/j.1521-0391.2003.tb00537.x

United States Census Bureau. (n.d.). *U.S. Census Bureau quickfacts: United States*. Census.gov. Retrieved April 10, 2023, from https://www.census.gov/quickfacts/fact/table/US/PST045221

Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

West, S. L., Gartlehner, G., & Mansfield, A. J. (2010). *Comparative effectiveness review methods: Clinical heterogeneity*. Agency for Healthcare Research and Quality.

Table 1. Descriptive Analysis of Primary Study Demographic and Contextual Factors

Demographic Variables	**K (29 total)**	**Mean (SD)**	**Percentage**
Mean years of education	14	12.94 (0.82)	
Age	28	37.67 (7.27)	
% Female	29		32.40%
% Transgender	3		19.68%
% Heterosexual	4		58.58%
% White	26		68.22%
% Black	21		36.86%
% Hispanic/Latino	15		11.55%
% Native American	6		4.31%
% Asian or NH/PI	6		1.51%
% more than one race	6		7.86%
% Some College	8		51.09%
% Less than 20k annual income*	2		56.15%
Context Variables:			
Study context:			
Specialty Treatment/SUD	16		55.17%
Community samples	9		31.03%
College setting	1		3.45%
Criminal justice	2		6.90%
Other setting	1		3.45%
Treatment funding source:			
Private	4		13.79%
Public	22		75.86%
Treatment setting:			
Outpatient	23		79.31%
Inpatient	3		10.34%
Other Variables:			
Drug Type:			
Alcohol	10		34.48%
Poly Drug	9		31.03%
Cannabis	3		10.34%
Other Drug	7		24.14%
Co-occurring disorder Study:			
Yes	12		41.38%
No	17		58.62%
Biological assay outcome measure	10		34.48%
Treatment length	29	13.67 (7.94)	
CBT format:			
Individual	15		51.72%
Group	13		44.83%
Mixed	1		3.45%
Cochrane Risk of Bias:			
Low risk	11		37.93%
Some Concerns	3		10.34%
High Risk	15		51.72%

Note. NH/PI = Native Hawaiian/Pacific Islander, SUD = Substance Use Disorder

Table 2. Pooled Effect Sizes for AOD Use Outcomes

Comparison	k	Std diff in means (95% CI)	z	p	I^2
Between group* early follow-up	29	d = 0.158 (95% CI = 0.079, 0.238)	3.911	0.000	46.296
Within group early follow-up	23	d_z = 1.147 (95% CI = 0.811, 1.482)	6.700	0.000	96.203

Note. *Late Follow-up from Thornton, 2003 included due to lack of early follow-up. Early follow-up was 1-6 months.

Table 3. Moderator Analysis at Early Follow-up by Comparison

	Between Group Early Follow-up					Within Group Early Follow-up				
Demographic Variables	k	β	95% CI	P	R^2 Analog	k	β	95% CI	P	R^2 Analog
Age	28	-0.0114	-0.0262, 0.0034	0.1313	0.00	22	-0.0248	-0.0727, 0.0232	0.3119	0.00
% Female	29	0.0030	-0.0017, 0.0078	0.2093	0.03	23	-0.0087	-0.0242, 0.0069	0.2762	0.03
% White	26	0.0025	-0.0005, 0.0054	0.1037	0.00	21	0.0133	-0.0010, 0.0276	0.0679	0.00
% Black	21	0.0004	-0.0021, 0.0028	0.7613	0.00	15	-0.0073	-0.0208, 0.0062	0.2913	0.00
% Hispanic/Latino	15	-0.0024	-0.0080, 0.0032	0.4069	0.00	12	-0.0322	-0.0884, 0.0239	0.2601	0.00
Mean years of education	14	-0.0312	-0.2134, 0.1510	0.7370	0.00	11	0.2389	-0.4839, 0.9617	0.5171	0.00
Context Variables:										
Study context	29	-0.1156	-0.2790, 0.0479	0.1657	0.00	23	-0.0798	-0.7528, 0.5932	0.8163	0.00
Treatment funding source	26	-0.2760	-0.5589, 0.0070	0.0559	0.26	21	0.7490	-0.3697, 1.8676	0.1894	0.00
Treatment setting	26	0.0037	-0.2173, 0.2248	0.9736	0.00	20	-0.4505	-1.6597, 0.7586	0.4652	0.00
Other Variables:										
Drug type	29	**0.1829***	0.0391, 0.3286	0.0127	0.37	23	0.0267	-0.6830, 0.7365	0.9411	0.00
Dual Dx Study	29	0.0528	-0.1201, 0.2257	0.5495	0.00	23	**-0.7276***	-1.4011, -0.0541	0.0342	0.00
Biological assay outcome measure	29	**-0.2356***	-0.4366, -0.0346	0.0216	0.34	23	0.6770	-0.0663, 1.4204	0.0743	0.00
Treatment length	29	-0.0034	-0.0150, 0.0082	0.5674	0.00	23	-0.0194	-0.0779, 0.0391	0.5158	0.00
Study Date	29	-0.0019	-0.0101, 0.0062	0.6455	0.00	23	**-0.0492***	-0.0792, -0.0193	0.0013	0.18
CBT format	28	-0.0311	-0.2086, 0.1465	0.7317	0.00	22	-0.6089	-1.3659, 0.1482	0.1150	0.00
Cochrane Risk of Bias	29	-0.0572	-0.2224, 0.1088	0.4976	0.00	23	0.3776	-0.3315, 1.0866	0.2966	0.00

Notes. * Statistically significant. Ref. group for study context was Specialty Tx/SUD. Ref. group for treatment funding source was private. Ref. group for treatment setting was inpatient. Ref. group for drug type was alcohol. Ref group for Dual Dx study was no. Ref group for CBT format was group. Ref group for biological assay outcome measure was yes.

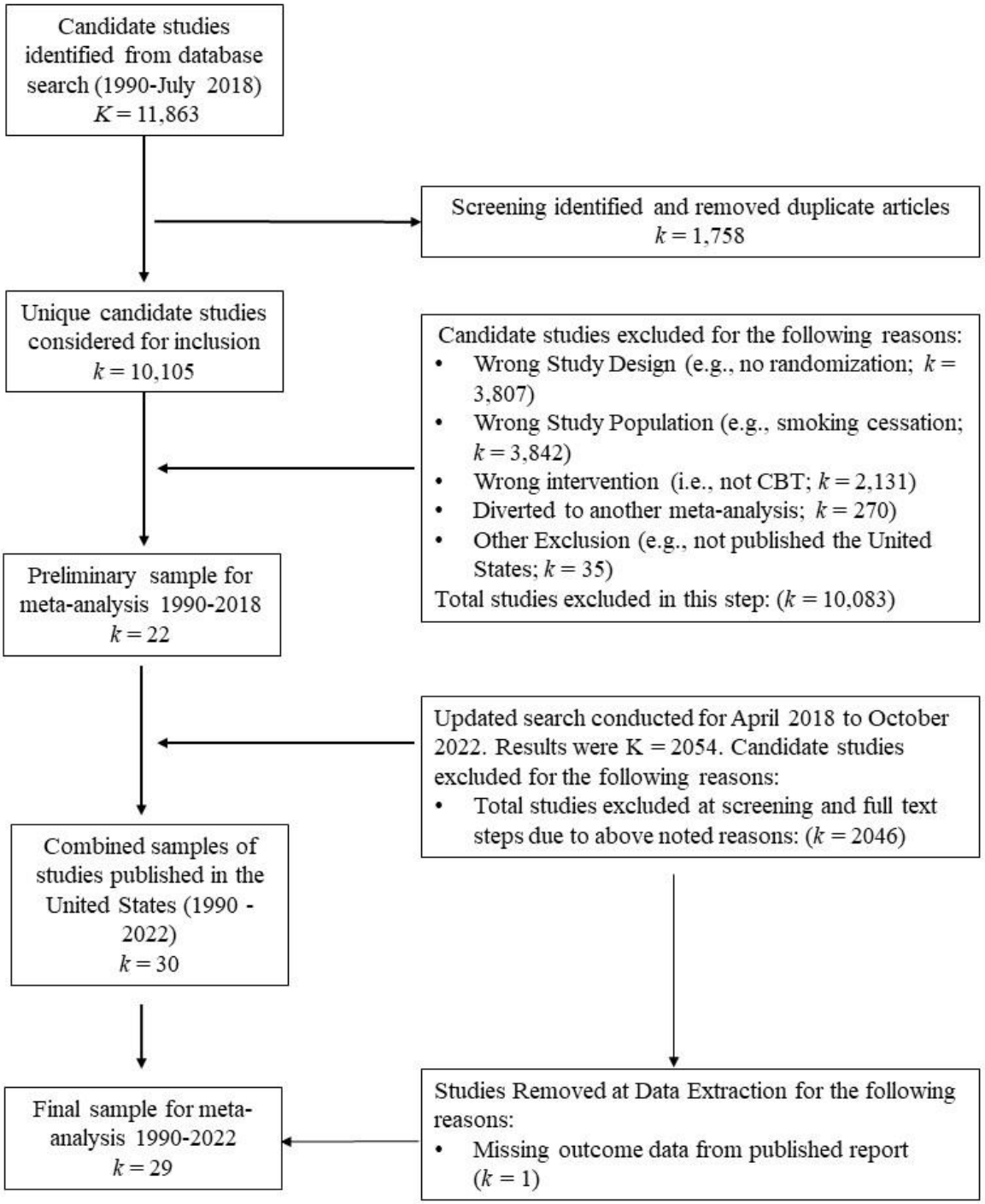


Figure 1. Flow of primary study inclusion.

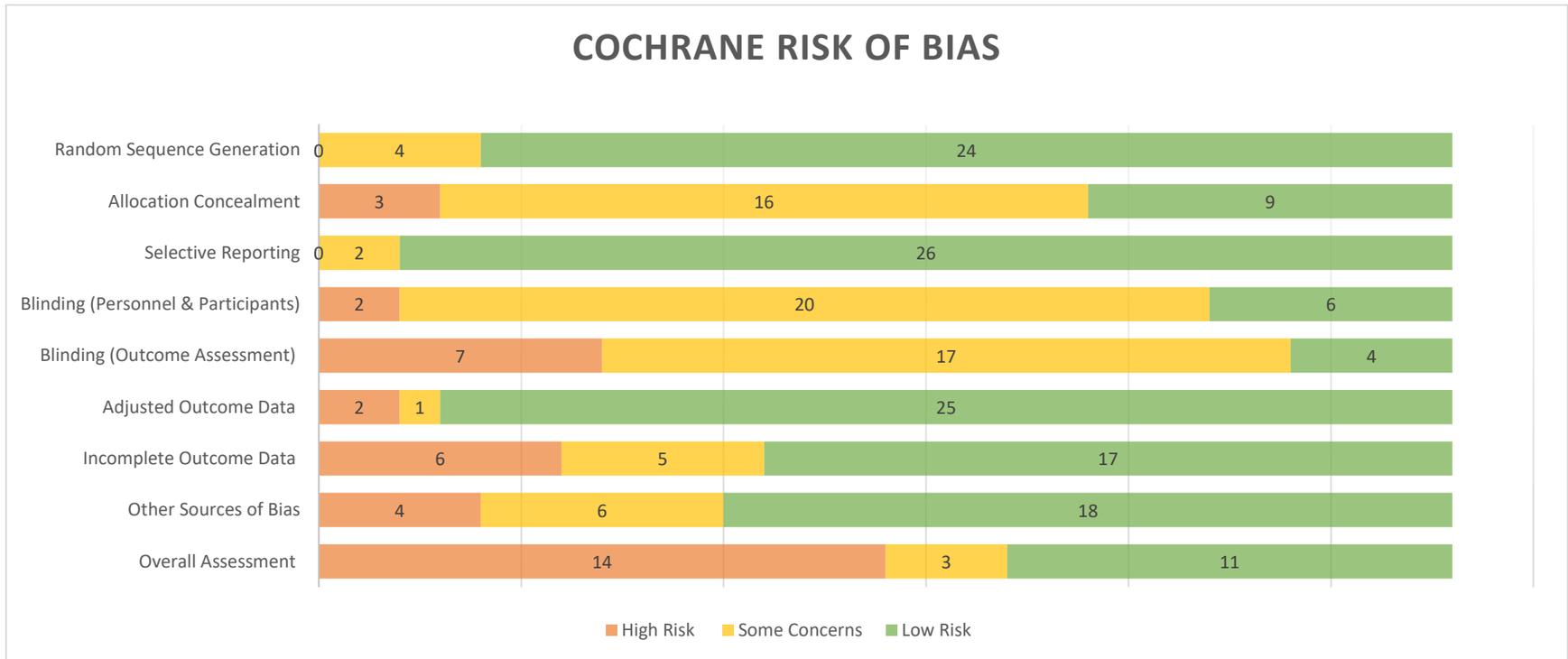


Figure 4. Distribution of risk of bias domain and overall ratings

Notes. *Random Sequence Generation* = Described the method used to generate the allocation sequence to allow an assess whether it should produce comparable groups. *Allocation Concealment* = Described the method used to conceal the allocation sequence to determine whether intervention allocations could have been foreseen before or during enrollment. *Selective Reporting* = Examined how the possibility of selective outcome reporting was examined by the authors and what was found. *Blinding (Personnel & Participants)* = Described all measures used, if any, to blind study participants and personnel from knowledge of which intervention a participant received. *Blinding (Outcome Assessment)* =Described all measures used, if any, to blind outcome assessors from knowledge of which intervention a participant received. *Adjusted Outcome Data* = Described whether outcome data used in analysis was adjusted for any reason. *Incomplete Outcome Data* = Described the completeness of outcome data for each main outcome, including attrition and exclusions from the analysis. *Other sources of Bias* = Includes any concerns about bias not addressed in other domains. *Overall Assessment* = This is determined qualitatively after looking at all domains.

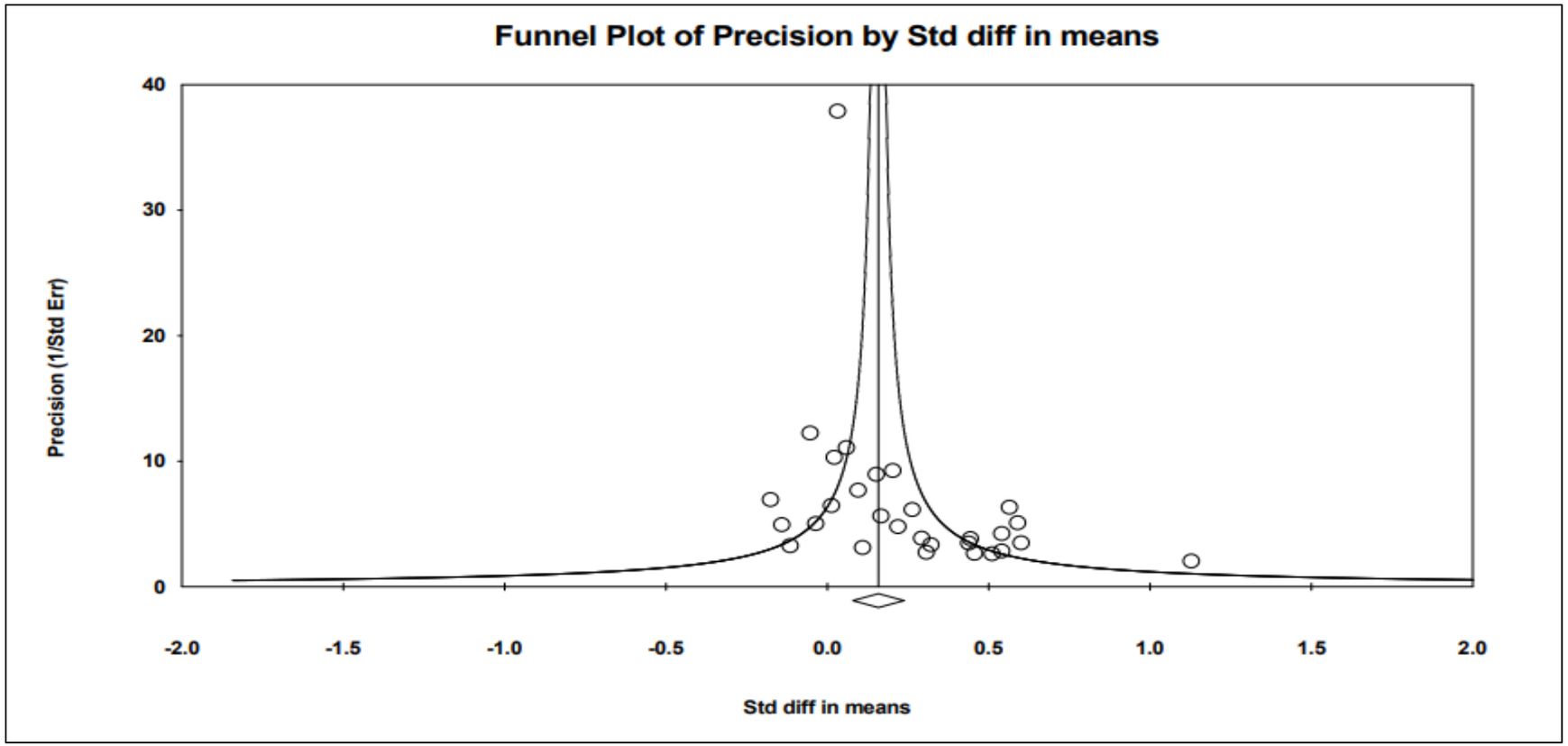


Figure 5. Plot of assessment of publication bias.
Notes. Assessment of bias in between-group CBT effect at early follow-up. Plot shows some asymmetry. The rank order correlation shows a positive and significant correlation between precision and effect size, indicating that smaller, less precise studies with larger effects may have been preferentially published ($\tau = .29$, $p = 0.027$).

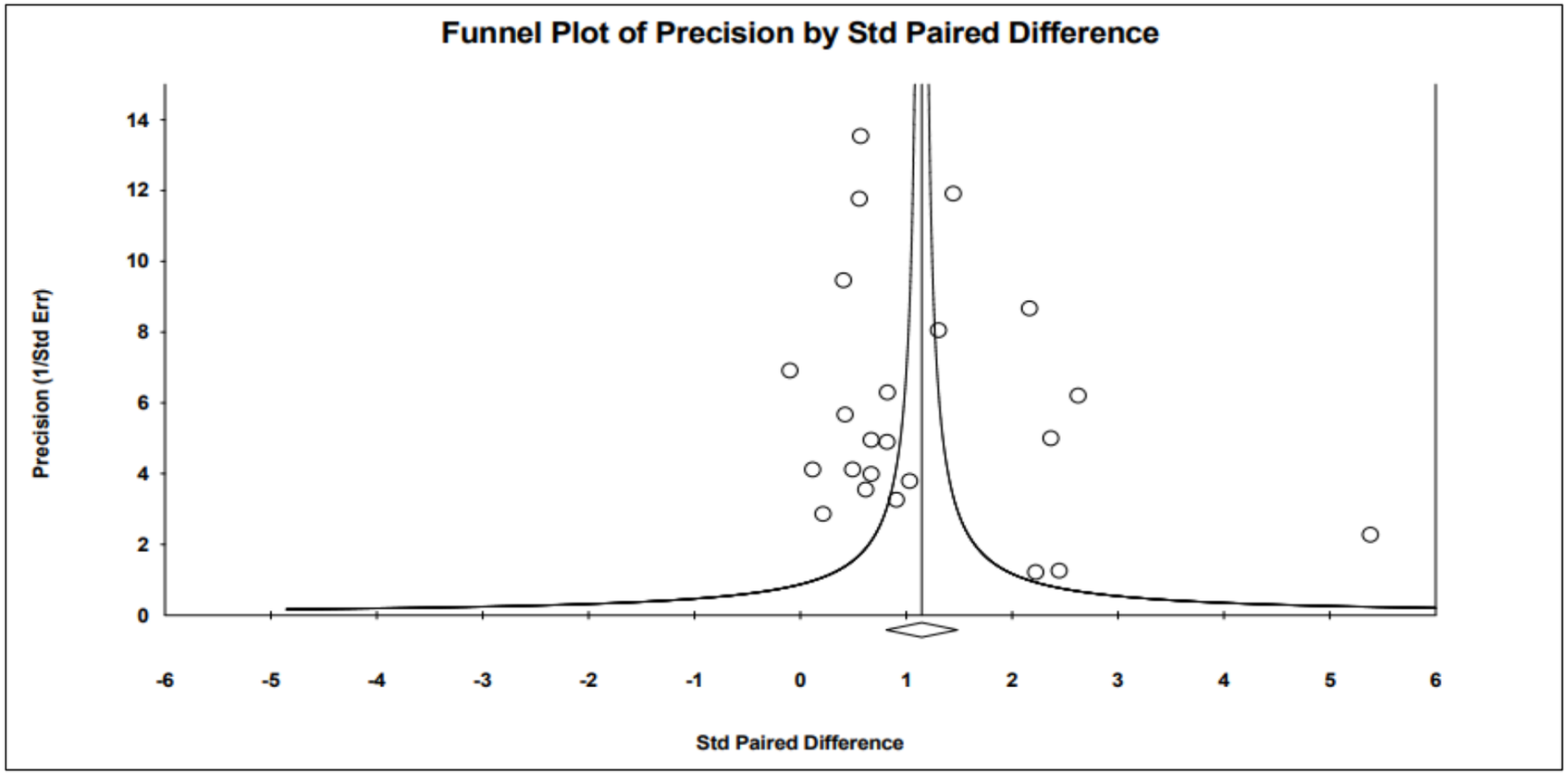
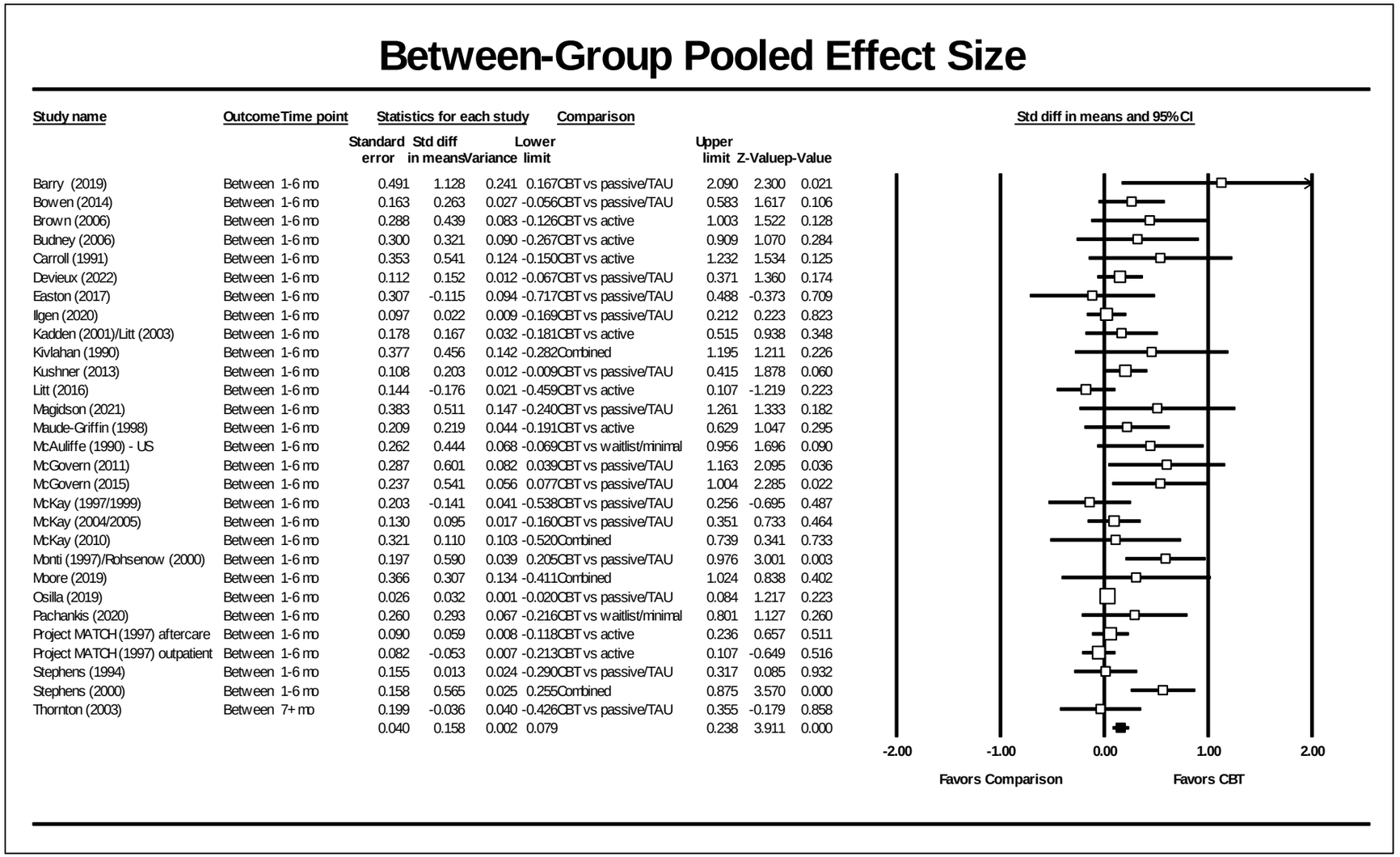


Figure 6. Plot of assessment of publication bias.
Notes. Assessment of bias in within-group CBT effect at early follow-up. Plot shows some asymmetry. The rank order correlation however shows a non-significant relationship between precision and effect size ($\tau = 0.16$, $p = 0.279$).



Study name	Outcome	Time point	Standard error	Std diff in means	Variance	Lower limit	Comparison	Upper limit	Z-Value	p-Value
Barry (2019)	Between	1-6 mo	0.491	1.128	0.241	0.167	CBT vs passive/TAU	2.090	2.300	0.021
Bowen (2014)	Between	1-6 mo	0.163	0.263	0.027	-0.056	CBT vs passive/TAU	0.583	1.617	0.106
Brown (2006)	Between	1-6 mo	0.288	0.439	0.083	-0.126	CBT vs active	1.003	1.522	0.128
Budney (2006)	Between	1-6 mo	0.300	0.321	0.090	-0.267	CBT vs active	0.909	1.070	0.284
Carroll (1991)	Between	1-6 mo	0.353	0.541	0.124	-0.150	CBT vs active	1.232	1.534	0.125
Devieux (2022)	Between	1-6 mo	0.112	0.152	0.012	-0.067	CBT vs passive/TAU	0.371	1.360	0.174
Easton (2017)	Between	1-6 mo	0.307	-0.115	0.094	-0.717	CBT vs passive/TAU	0.488	-0.373	0.709
Ilgen (2020)	Between	1-6 mo	0.097	0.022	0.009	-0.169	CBT vs passive/TAU	0.212	0.223	0.823
Kadden (2001)/Litt (2003)	Between	1-6 mo	0.178	0.167	0.032	-0.181	CBT vs active	0.515	0.938	0.348
Kivlahan (1990)	Between	1-6 mo	0.377	0.456	0.142	-0.282	Combined	1.195	1.211	0.226
Kushner (2013)	Between	1-6 mo	0.108	0.203	0.012	-0.009	CBT vs passive/TAU	0.415	1.878	0.060
Litt (2016)	Between	1-6 mo	0.144	-0.176	0.021	-0.459	CBT vs active	0.107	-1.219	0.223
Magidson (2021)	Between	1-6 mo	0.383	0.511	0.147	-0.240	CBT vs passive/TAU	1.261	1.333	0.182
Maude-Griffin (1998)	Between	1-6 mo	0.209	0.219	0.044	-0.191	CBT vs active	0.629	1.047	0.295
McAuliffe (1990) - US	Between	1-6 mo	0.262	0.444	0.068	-0.069	CBT vs waitlist/minimal	0.956	1.696	0.090
McGovern (2011)	Between	1-6 mo	0.287	0.601	0.082	0.039	CBT vs passive/TAU	1.163	2.095	0.036
McGovern (2015)	Between	1-6 mo	0.237	0.541	0.056	0.077	CBT vs passive/TAU	1.004	2.285	0.022
McKay (1997/1999)	Between	1-6 mo	0.203	-0.141	0.041	-0.538	CBT vs passive/TAU	0.256	-0.695	0.487
McKay (2004/2005)	Between	1-6 mo	0.130	0.095	0.017	-0.160	CBT vs passive/TAU	0.351	0.733	0.464
McKay (2010)	Between	1-6 mo	0.321	0.110	0.103	-0.520	Combined	0.739	0.341	0.733
Monti (1997)/Rohsenow (2000)	Between	1-6 mo	0.197	0.590	0.039	0.205	CBT vs passive/TAU	0.976	3.001	0.003
Moore (2019)	Between	1-6 mo	0.366	0.307	0.134	-0.411	Combined	1.024	0.838	0.402
Osilla (2019)	Between	1-6 mo	0.026	0.032	0.001	-0.020	CBT vs passive/TAU	0.084	1.217	0.223
Pachankis (2020)	Between	1-6 mo	0.260	0.293	0.067	-0.216	CBT vs waitlist/minimal	0.801	1.127	0.260
Project MATCH (1997) aftercare	Between	1-6 mo	0.090	0.059	0.008	-0.118	CBT vs active	0.236	0.657	0.511
Project MATCH (1997) outpatient	Between	1-6 mo	0.082	-0.053	0.007	-0.213	CBT vs active	0.107	-0.649	0.516
Stephens (1994)	Between	1-6 mo	0.155	0.013	0.024	-0.290	CBT vs passive/TAU	0.317	0.085	0.932
Stephens (2000)	Between	1-6 mo	0.158	0.565	0.025	0.255	Combined	0.875	3.570	0.000
Thornton (2003)	Between	7+ mo	0.199	-0.036	0.040	-0.426	CBT vs passive/TAU	0.355	-0.179	0.858
			0.040	0.158	0.002	0.079		0.238	3.911	0.000

Figure 7. Forrest plot showing between-group pooled effect size.

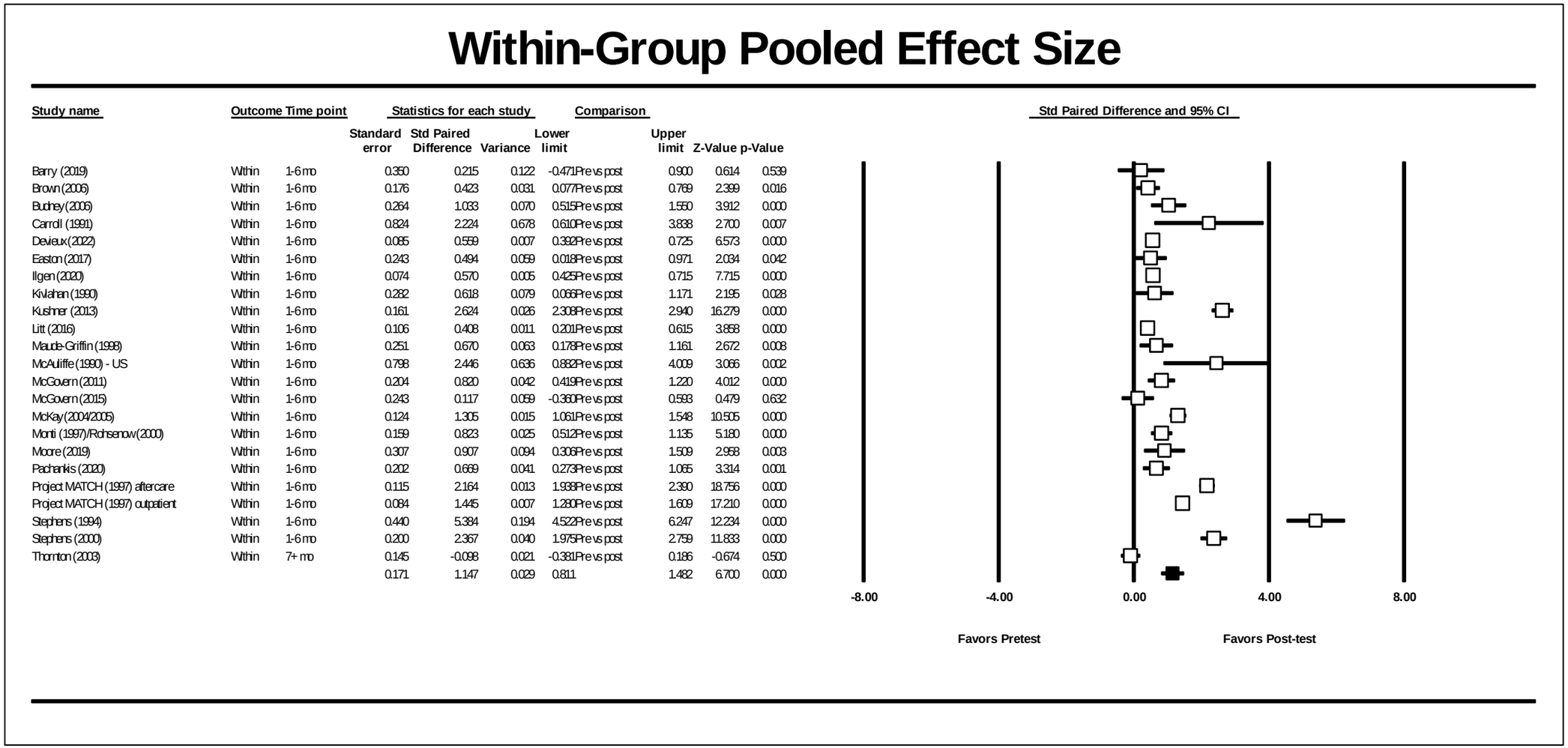


Figure 8. Forrest plot showing between-group pooled effect size.

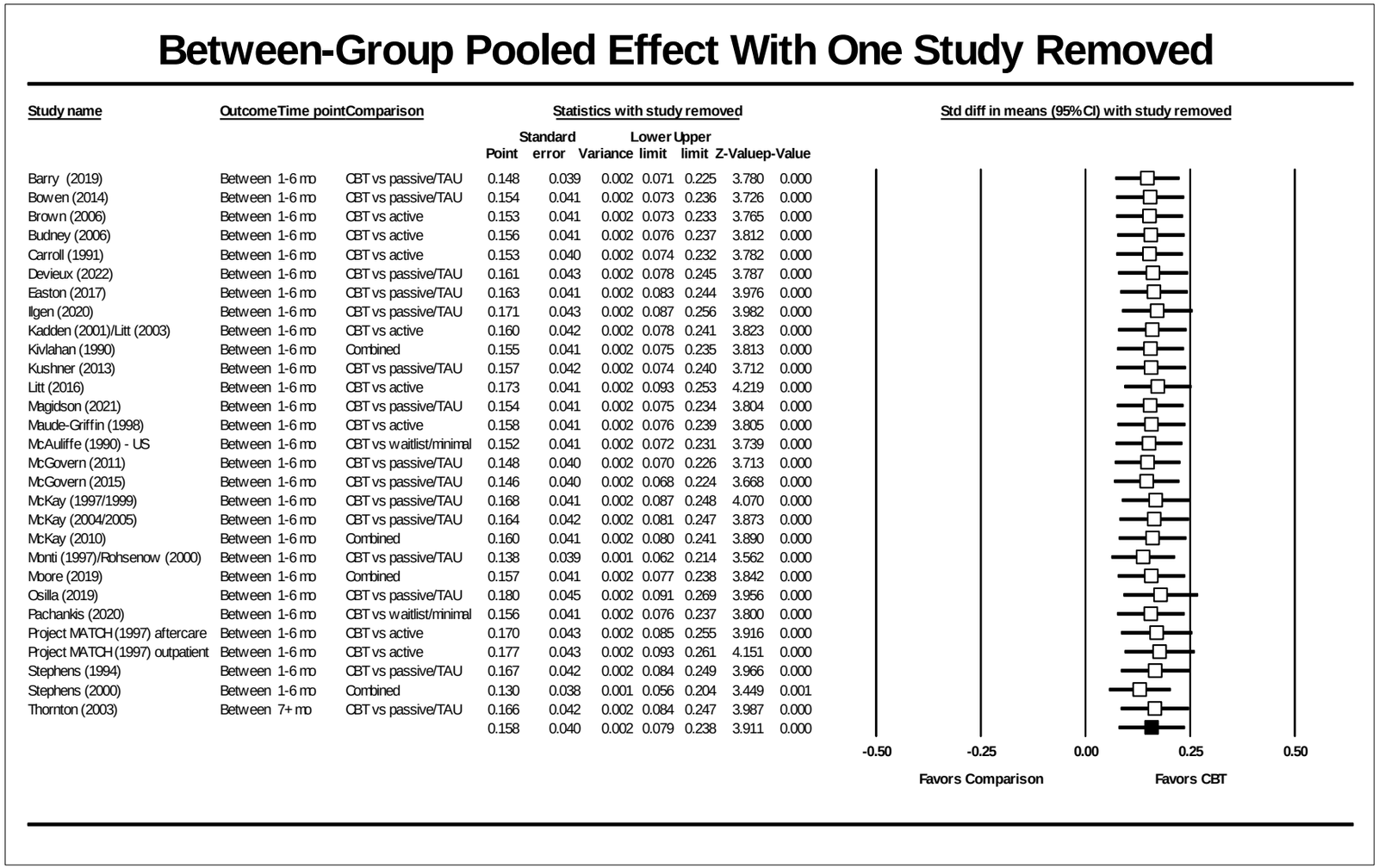


Figure 9. Forrest plot showing between-group pooled effect with one study removed.
Notes. Based on plot no one study would significantly change the effect size if removed.

Within-Group Pooled Effect With One Study Removed

Study name	Comparison	Outcome	Time point	Statistics with study removed						
				Point	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Barry (2019)	Pre vs post	Within	1-6 mo	1.186	0.175	0.031	0.843	1.529	6.781	0.000
Brown (2006)	Pre vs post	Within	1-6 mo	1.183	0.177	0.031	0.836	1.530	6.685	0.000
Budney (2006)	Pre vs post	Within	1-6 mo	1.153	0.176	0.031	0.807	1.499	6.532	0.000
Carroll (1991)	Pre vs post	Within	1-6 mo	1.121	0.173	0.030	0.782	1.461	6.479	0.000
Devieux (2022)	Pre vs post	Within	1-6 mo	1.181	0.184	0.034	0.821	1.542	6.421	0.000
Easton (2017)	Pre vs post	Within	1-6 mo	1.178	0.176	0.031	0.832	1.523	6.683	0.000
Ilgen (2020)	Pre vs post	Within	1-6 mo	1.182	0.186	0.035	0.817	1.547	6.350	0.000
Kivlahan (1990)	Pre vs post	Within	1-6 mo	1.171	0.176	0.031	0.826	1.516	6.652	0.000
Kushner (2013)	Pre vs post	Within	1-6 mo	1.065	0.162	0.026	0.747	1.383	6.562	0.000
Litt (2016)	Pre vs post	Within	1-6 mo	1.186	0.179	0.032	0.835	1.537	6.621	0.000
Maude-Griffin (1998)	Pre vs post	Within	1-6 mo	1.170	0.176	0.031	0.824	1.515	6.628	0.000
McAuliffe (1990) - US	Pre vs post	Within	1-6 mo	1.115	0.173	0.030	0.776	1.454	6.445	0.000
McGovern (2011)	Pre vs post	Within	1-6 mo	1.164	0.177	0.032	0.816	1.512	6.556	0.000
McGovern (2015)	Pre vs post	Within	1-6 mo	1.195	0.175	0.031	0.852	1.537	6.827	0.000
McKay (2004/2005)	Pre vs post	Within	1-6 mo	1.142	0.180	0.033	0.788	1.495	6.329	0.000
Monti (1997)/Rohsenow(2000)	Pre vs post	Within	1-6 mo	1.165	0.179	0.032	0.814	1.516	6.507	0.000
Moore (2019)	Pre vs post	Within	1-6 mo	1.158	0.176	0.031	0.813	1.503	6.579	0.000
Pachankis (2020)	Pre vs post	Within	1-6 mo	1.171	0.177	0.031	0.823	1.518	6.604	0.000
Project MATCH (1997) aftercare	Pre vs post	Within	1-6 mo	1.088	0.164	0.027	0.766	1.410	6.617	0.000
Project MATCH (1997) outpatient	Pre vs post	Within	1-6 mo	1.136	0.183	0.033	0.778	1.494	6.217	0.000
Stephens (1994)	Pre vs post	Within	1-6 mo	0.980	0.160	0.026	0.667	1.293	6.132	0.000
Stephens (2000)	Pre vs post	Within	1-6 mo	1.085	0.170	0.029	0.751	1.418	6.371	0.000
Thornton (2003)	Pre vs post	Within	7+ mo	1.206	0.171	0.029	0.870	1.542	7.039	0.000
				1.147	0.171	0.029	0.811	1.482	6.700	0.000

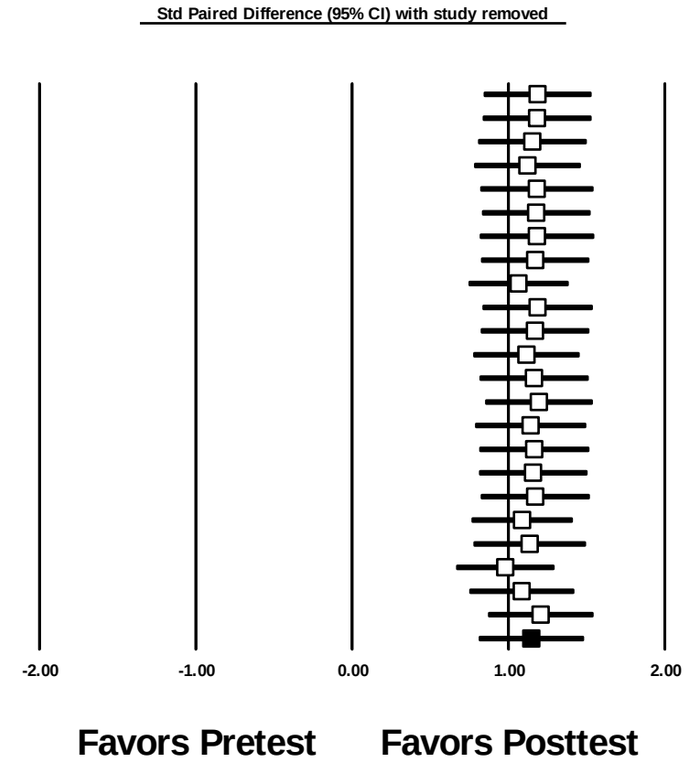


Figure 10. Forrest plot showing between-group pooled effect with one study removed.
Notes. Based on plot no one study would significantly change the effect size if removed.

Appendix A

PubMed Search Strategy

((("Cognitive Behavioral Therapy"[MeSH Terms] OR "Secondary Prevention"[MeSH Terms] OR "Behavior Therapy"[MeSH Terms:noexp] OR "Behavior Therapy"[Text Word] OR "behaviour therapy"[Text Word] OR "behavioral therapy"[Text Word] OR "behavioral therapies"[Text Word] OR "behavioural therapy"[Text Word] OR "behavioural therapies"[Text Word] OR "CBI"[Text Word] OR "CBT"[Text Word] OR "cognitive therapy"[Text Word] OR "cognitive therapies"[Text Word] OR "cognitive behavioral"[Text Word] OR "cognitive behavioural"[Text Word] OR ("relapse prevention"[Text Word] OR "Secondary Prevention"[Text Word]) OR ("coping skills"[Text Word] OR "coping behavio*"[Text Word] OR "social skills training"[Text Word])) AND ("Substance-Related Disorders"[MeSH Terms] OR "Illicit Drugs"[MeSH Terms] OR "Designer Drugs"[MeSH Terms] OR "Heroin"[MeSH Terms] OR ("alcohol"[Text Word] OR "cocaine"[Text Word] OR "methamphetamine"[Text Word] OR "stimulant"[Text Word] OR "opiate"[Text Word] OR "opioid"[Text Word] OR "Heroin"[Text Word] OR "marijuana"[Text Word] OR "cannabis"[Text Word] OR "illicit drug*"[Text Word] OR "illegal drug*"[Text Word] OR "polysubstance"[Text Word])) AND (("randomized controlled trial"[Publication Type] OR "controlled clinical trial"[Publication Type] OR "randomized"[Title/Abstract] OR "placebo"[Title/Abstract] OR "clinical trials as topic"[MeSH Terms] OR ("randomly"[Title/Abstract] OR "trial"[Title])) NOT ("animals"[MeSH Terms] NOT "humans"[MeSH Terms]))) AND (2018/7/1:3000/12/12[pdat])