An Evaluation of Cognitive Behavioral Therapy for Substance Use: An Application of Tolin's Criteria for Empirically Supported Treatments

Authors: Cassandra L. Boness, Victoria R. Votaw, Frank J. Schwebel,
David I.K. Moniz-Lewis, R. Kathryn McHugh, & Katie Witkiewitz

Tolin Criteria Report Submission Checklist

Checklist Item	Required Content	Reported on Page(s)	Reviewer 1	Reviewer 2	Consensus
Step 1. Treatment Nomination and Committee Formation					
1	Submit letter of intent to the Committee on Science and Practice (CSP); see Appendix C	3			
2	Document selection process of workgroup members	3			
3	Include a conflict of interest declaration from each member, noting the relationship, financial or otherwise, with the treatment being evaluated	3			
Step 2. Locating and Screening Reviews for Inclusion					
4	Document search terms	3-4			
5	Document databases searched	4			
6	Document inclusion and exclusion criteria	4			
7	Document the rationale and decision regarding restricting the search based on publication date if applicable	4			
8	Flow diagram (or comparable) documenting screening and selection of final reviews	6			
9	PICOTS table	7-8			
Step 3. Assessment of Review Quality					
10	Describe critical domains	8			
11	AMSTAR 2 (or comparable) table	9			
12	AMSTAR 2 (or comparable) text including description of how “low” or “critically low” studies will be handled	8			
Step 4. Evaluating Outcomes and Judging Quality of the Evidence					
13	Compile effect size values	10			
14	If relevant, statistically combine effect sizes and report accompanying measures of central tendency	10			
15	Forest plot (or comparable) diagram	11			
16	Narrative description of results	10			
17	Details of sensitivity analysis	12			
18	Complete “Quality of the Evidence” table	14			
Step 5. Consideration of Additional Contextual Factors					
19	Table documenting presence or absence of contextual factors	17			
Step 6. Overall Treatment Recommendation					
20	GRADE table with rating	18			
21	Narrative treatment recommendation, including what would be needed to improve the recommendation in the future	18			
IN THE CASE OF AN AMENDED REPORT					
22	Summary statement of updates				

Evaluation of Cognitive Behavioral Therapy for Substance Use

Committee: Cassandra L. Boness, Victoria R. Votaw, David I.K. Moniz-Lewis, Frank J. Schwebel, R. Kathryn McHugh, Katie Witkiewitz
Dates of Review: October 1, 2021 to January 13, 2022
Online Recommendation: Strong
Recommended Citation: Boness, C. L., Votaw, V., Schwebel, F. J., Moniz-Lewis, D. I., McHugh, R. K., & Witkiewitz, K. (2022, January 13). An Evaluation of Cognitive Behavioral Therapy for Substance Use: An Application of Tolin's Criteria for Empirically Supported Treatments. Retrieved from osf.io/rbx8s

The following document outlines the evaluation of cognitive behavioral therapy (CBT) for substance use according to the Tolin et al. (2015) criteria. This is part of a larger effort to revise and update the APA Division 12 ratings of the strength of psychological treatments for given diagnoses.

1. Treatment Nomination and Committee Formation

1.1 Nomination and Letter of Intent

The authors of the current evaluation submitted a letter of interest to the Division 12 Committee on Science and Practice to evaluate CBT for substance use disorder on September 23, 2021. The LOI was approved on October 1, 2021.

1.2 Selection Process for Treatment Evaluation Committee

When forming our committee, we considered individuals who have familiarity with cognitive and behavioral treatments for substance use disorder. Authors have expertise in systematic reviews (CLB, VRV, RKM), CBT (RKM, KW), and/or substance use (all authors).

1.3 Conflict of Interest Declaration

For each author, please describe potential conflicts of interest or certify that each author has no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria, educational grants, participation in speakers' bureaus, membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this Tolin Criteria Report.

Cassandra Boness: Dr. Boness is a member of the Division 12 Committee on Science and Practice and played a key role in the development of the "Tolin Criteria" manual. To mitigate this conflict, Dr. Boness was not involved in the evaluation or discussion of this evaluation report.
Victoria Votaw: No conflicts of interest to declare.
Frank Schwebel: No conflicts of interest to declare.
David Moniz-Lewis: No conflicts of interest to declare.
R. Kathryn McHugh: No conflicts of interest to declare.
Katie Witkiewitz: Dr. Witkiewitz was the primary author of a randomized controlled trial of relapse prevention, a form of CBT for substance use, which is a treatment included in our evaluation:

Witkiewitz, K., Warner, K., Sully, B., Barricks, A., Stauffer, C., Steckler, G., Thompson, B., & Luoma, J. (2014). Randomized trial comparing mindfulness-based relapse prevention with relapse prevention for women offenders at a residential addiction treatment center. Substance Use and Misuse, 49, 536-546. doi: 10.3109/10826084.2013.856922

To mitigate any potential impacts of Dr. Witkiewitz's conflict of interest, she has not contributed to any coding of the included quantitative reviews.

2. Locating and Screening Reviews for Inclusion

2.1 Search Terms

Search terms included the following: (TI ((“Cognitive Behavior* Therapy” or “CBT” or "Relapse prevention" or "coping skills training" or “skills training”) AND (addiction or alcohol or drug or substance or abuse or

dependence or cocaine or opioid or cannabis or marijuana or heroin or cocaine or amphetamine or prescription drug))) AND ((TI (Review or “Systematic Review” or “Quantitative Review" or “meta analysis”) OR (SU(Review or “Systematic Review” or “Quantitative Review” or “meta analysis”))) NOT TI ("qualitative review" or "narrative review") where TI = title and SU = subject. The search was restricted to the University of Missouri’s library collections. The search was replicated using the University of New Mexico’s library collections which resulted in a lower number of records; thus, we used the University of Missouri library collection to be as comprehensive as possible. Initial search results returned 47 records. The search strategy was conducted on September 24, 2021.

2.2 Databases

We searched 7 databases including Scopus, MEDLINE, ScienceDirect, PsycINFO, Social Sciences Citation Index, Science Citation Index, and Academic Search Premier.

2.3 Inclusion and Exclusion Criteria

For the purposes of the current synthesis, we limited our eligible reviews to quantitative meta-analyses only. This is consistent with Tolin et al. (2015) which proposes that candidate treatments be evaluated on the basis of existing quantitative reviews.

We define CBT as a multisession intervention that targets cognitive, affective, behavioral, and/or environmental risks for substance use and provides training in skills to help an individual achieve and maintain substance use abstinence or moderation or reduce harm related to substance use. Consistent with prior meta-analyses in the areas of CBT for substance use (e.g., Magill et al., 2019), we considered Relapse Prevention (RP) and coping skills training cognitive-behavioral interventions, given such treatments use critical CBT elements, such as functional analysis, avoidance of high-risk situations, and drug refusal skills, among others.

Inclusion criteria included being a quantitative review focused on CBT for substance use disorder (SUD), substance use, or substance-related problems among adults. Exclusion criteria included lack of reporting on substance use-related outcomes (e.g., quantity, frequency of use) and a sole focus on CBT + pharmacotherapy. Notably, we excluded quantitative reviews that only focused solely on CBT for nicotine dependence, given that meta-analyses examining CBT for substance use have generally excluded trials focused on nicotine dependence (e.g., Dutra et al., 2008; Magill et al., 2019) to limit the scope and draw targeted conclusions. We also excluded quantitative reviews focused only on Behavioral Couples Therapy. Although Behavioral Couples Therapy is a form of CBT, modules administered and hypothesized mechanisms of change differ substantially between traditional (individual or group) CBT and Behavioral Couples Therapy (McHugh, Hearon, & Otto, 2010).

2.4 Dates of Publication

There were no restrictions on language or year of publication.

2.5 Selection of Reviews to Include in the Evaluation

Our search strategy returned 47 results. After eliminating duplicates there were 44 possibly eligible records for inclusion (43 unique records and one erratum). Reviewers double coded each of the 44 meta-analyses as eligible, not eligible, or possibly eligible based on their title. For meta-analyses coded as eligible or possibly eligible, full texts were obtained and read to further determine eligibility. All discrepancies were resolved via consensus among all committee members. Of the 44 possibly eligible records, 12 were excluded because they were not focused on CBT for substance use as a standalone intervention (e.g., they focused on other interventions for substance use or primarily on pharmacotherapy or CBT + pharmacotherapy), 22 were excluded for not being a meta-analysis, and 12 were excluded for not reporting on substance-related outcomes (e.g., they primarily enrolled individuals with behavioral addictions or other psychological disorders [e.g., depression, trauma] and therefore did not report on substance-related outcomes). This resulted in two eligible meta-analyses. We also conducted backwards searches of these meta-analyses which resulted in 7 additional possibly eligible manuscripts. Upon full-text review, 3 of these were eligible. This resulted in 5 eligible meta-analyses for inclusion in the current review (see Figure 1).

For each of the 5 eligible meta-analyses, two members (CB and VV) double coded the PICOTS (population, intervention, comparison, outcomes, timeline, setting) criteria (e.g., Schardt et al., 2007). PICOTS allows for a full consideration of review characteristics (see Table 1) and assists the reader in evaluating the appropriateness of the eligible reviews included for answering the clinical question of interest. Discrepancies among coders were resolved by consensus between CB and VV.

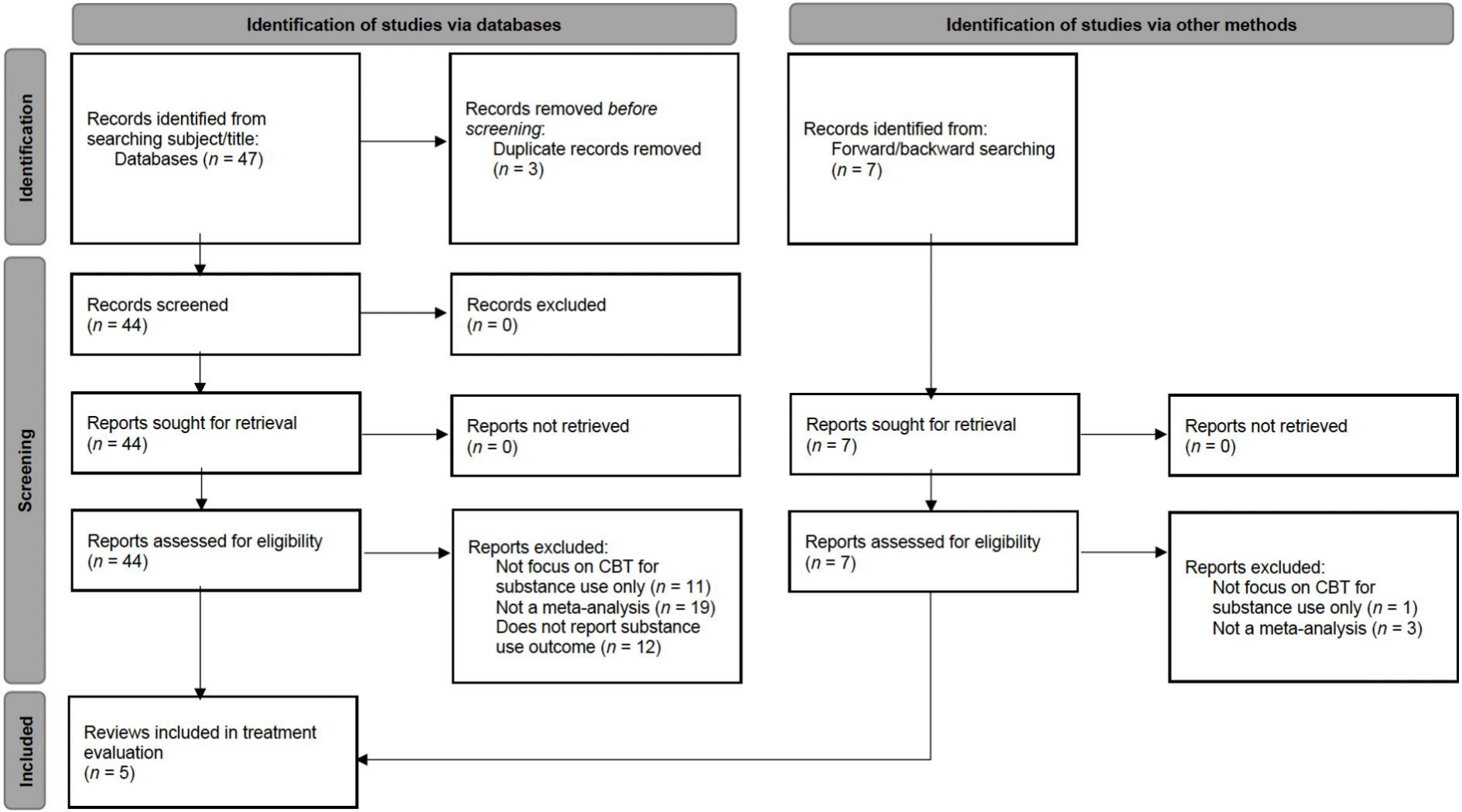

Figure 1. Flow Chart for Cognitive Behavioral Therapy for Substance Use Search Process. This figure illustrates the search process for locating reviews eligible for inclusion in the treatment evaluation.

2.6 PICOTS

Table 1

Description of Reviews Included in the CBT for SUD Treatment Evaluation

Study	Intervention(s)	Number of Studies	Population	Setting	Comparison Condition	Outcome	Time Points
Dutra et al., 2008	Psychosocial treatments for SUD, including CBT and RP	18 (13 CBT; 5 RP)	Adults with illicit SUDs, including cannabis (*n* = 4), cocaine (*n* = 4), opioids (*n* = 4), polysubstance (*n* = 6)	Non-intensive outpatient treatment	Inactive or active treatment	Self-reported or biologically-verified substance use	Post-treatment
Irvin et al., 1999	RP (individual, group, or couples format; with or without an adjunctive treatment)	26	Alcohol (*n* = 10), smoking (*n* = 8), or other substance use, including polysubstance use (*n* = 5) and cocaine use (*n* = 3)	Inpatient; outpatient	Pre-post change within person; waitlist or no-additional treatment control; active intervention	Self-reported or biologically-verified substance use; psychosocial adjustment	Post-treatment; 1-month, 3-month, 6-month, and 12-months post-treatment
Magill & Ray, 2009	CBT, RP, or coping-skills training (individual or group format; with or without an adjunctive treatment)	53	Adults with SUDs, including alcohol (*n* = 23), cocaine/stimulants (*n* = 11), polydrug (*n* = 11), cannabis (*n* = 6), opioids (*n* = 2)	NS	Active treatment; passive treatment or usual service; no treatment; no CBT adjunct	Self-reported or biologically-verified substance use	Post-treatment to 4 months post-treatment; 6-12 months post-treatment
Magill et al. 2019	CBT or RP (individual or group format)	30 (across 32 study sites)	Adults with SUDs or problematic use, including alcohol (*n* = 15), cannabis (*n* = 3), opioids (*n* = 2), stimulants (*n* = 6), and polydrug (*n* = 6)	Community sample; specialty substance use or mental health clinic; Medical setting; College setting; Criminal justice setting; Other setting	Minimal treatment; Non-specific therapy; Other specific therapy	Self-reported or biologically-verified substance use	Early (1-6 months) post-treatment; Late post-treatment (8+ months)
Windsor et al., 2015	CBT (individual, group, or combined)	16	Adults (≥70% White or Black and/or Hispanic) with substance use, including cocaine (*n* = 3), cannabis (*n* = 3), alcohol (*n* = 6),	NS	Comparison treatment; pre-post change within person	Substance use	Post-treatment; Average follow-up from baseline (1-24 months)

	alcohol + other drugs (n = 4)

Note. CBT = Cognitive Behavioral Therapy; RP = Relapse Prevention; SUD = Substance Use Disorder; NS = not specified.

3. Assessment of Review Quality

For each meta-analysis, committee members (DM and CB) double coded the AMSTAR2 (Shea et al., 2017) items. Discrepancies among coders were resolved between DM, CB, and VV. Those that are rated as low or critically low (n = 4) were excluded from the main effect size estimates but were considered in supplemental sensitivity analyses (below). Of note, we did not separately assess treatment fidelity or risk of bias given similar content was assessed by the AMSTAR2 items.

3.1 Identifying Critical Domains

Consistent with the CBT-I evaluation (Boness et al., 2020), we chose the following six domains as "critical" (see also bolded items in Table 2): including components of PICOTS in the research question and inclusion criteria, using a comprehensive literature search strategy, providing an adequate description of included studies, using appropriate methods for the statistical combination of results (i.e., estimating a combined effect size), accounting for risk of bias in the primary studies included when interpreting results, and providing satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review.

3.2 Assigning an Overall Confidence Rating

Independent raters indicated their overall confidence (*critically low*, *low*, *moderate*, and *high*) in the results of each meta-analysis based on the pattern of results of AMSTAR2, including consideration of critical and non-critical domains.

3.3 Managing Reviews with Poor Quality

Given four of the five meta-analyses considered eligible for inclusion in the current evaluation were of low or critically low quality, the remaining steps of the evaluation solely consider data from Magill et al., 2019. However, to be thorough, we also include supplemental sensitivity analyses that consider all reviews regardless of quality. This is consistent with the Tolin Criteria guidance document (Boness et al., *unpublished*).

Table 2

AMSTAR2 Results for Eligible Studies

Item	Irvin et al. (1999)	Dutra et al. (2008)	Magill & Ray (2009)	Windsor et al. (2015)	Magill et al. (2019)
1. Did the research questions and inclusion criteria for the review include the components of PICO?	Y	Y	Y	Y	Y
2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?	N	N	N	N	N
3. Did the review authors explain their selection of the study designs for inclusion in the review?	N	N	Y	Y	N
4. Did the review authors use a comprehensive literature search strategy?	N	N	PY	N	PY
5. Did the review authors perform study selection in duplicate?	N	N	N	Y	Y
6. Did the review authors perform data extraction in duplicate?	N	N	N	Y	Y
7. Did the review authors provide a list of excluded studies and justify the exclusions?	N	N	N	N	N
8. Did the review authors describe the included studies in adequate detail?	N	PY	PY	PY	PY
9. Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?					
RCT	N	N	N	N	Y
NRSI	N	NA	NA	NA	NA
10. Did the review authors report on the sources of funding for the studies included in the review?	N	N	N	N	N
11. If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?					
RCT	N	N	Y	Y	Y
NRSI	Y	NA	NA	NA	NA
12. If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	N	N	N	N	Y
13. Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?	N	N	N	N	Y
14. Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	N	Y	Y	Y	Y
15. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?	Y	Y	Y	Y	Y
16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	Y	N	Y	Y	Y
Overall Rating	Critically low	Critically low	Low	Critically low	Moderate

Note. Y = yes, PY = partial yes, N = no, NA = not applicable, PICO: (P = population, I = intervention, C = comparator group, O = outcome), RCT = randomized controlled trial, NRSI = non-randomized studies of interventions. Items in bold are considered critical weaknesses if coded "no." All studies were double coded and discrepancies were resolved by consensus to arrive at the final ratings.

4. Evaluating Outcomes of Review and Judging the Quality of the Evidence

4.1 Creating Procedures for Extracting Data

Effect sizes were extracted for all meta-analyses regardless of quality. VV and FS each listed descriptions for effect sizes of interest (e.g., by comparison group, time point, substance) from each of the five meta-analyses. CB was responsible for comparing these descriptions and resolving discrepancies. CB then developed an excel spreadsheet to be used separately by VV and FS in effect size (and confidence interval) extraction.

4.2 Data Collection and Validation

VV and FS separately extracted effect sizes and corresponding confidence intervals in their original metrics for all agreed upon outcomes of interest. CB then compared extracted data across the two excel sheets and resolved discrepancies.

4.3 Statistically Combining Effect Sizes

Because only a single meta-analysis (Magill et al., 2019) had sufficient quality for inclusion, per AMSTAR2, effect sizes were not statistically aggregated across separate meta-analyses. However, because Magill et al., 2019 reported four outcomes (early and late follow-up substance use frequency and early and late follow-up substance use quantity) for three separate comparator groups (minimal treatment [e.g., waitlist, brief psychoeducation], non-specific therapy [e.g., treatment as usual, supportive therapy, drug counseling], and specific therapy [e.g., Motivational Interviewing, Contingency Management]), we report two separate sets of effect sizes. First, we report on all comparator groups for each outcome separately. Second, we report on averaged effect sizes across the non-specific and specific comparator groups for each of the three outcomes to get a sense of how effect sizes differ for inactive versus active treatments. In this second set of effect sizes, we also averaged across substance use quantity and frequency to indicate overall substance use effect sizes. All estimates used Hedge's *g* and we considered Cohen's (1988) guidelines for classifying small effects as $g = 0.2$, medium effects as $g = 0.5$, and large effects as $g = 0.8$.

4.4 Interpreting Results

First, we report on the effect sizes by comparator group for each of the four outcomes (early and late follow-up substance use frequency and quantity). These estimates are displayed in Figure 2 and all raw effect size estimates are provided in Supplemental Table 1. Of note, for the late follow-up substance use quantity outcome, the only effect size reported was for the specific treatment comparator group, given these data were not reported in trials examining minimal and non-specific comparator groups.

Regarding the overall pattern of results, CBT was most effective at early follow-up (defined as 1-6 months post-treatment) compared to late follow-up (defined as 8+ months post-treatment) and had the largest effects when compared to minimal (inactive) treatment. Effect sizes for substance use quantity were generally larger than substance use frequency. In the minimal comparator group, effect sizes for early follow-up substance use frequency and quantity were medium in magnitude (>0.50) and the effect size for late follow-up frequency was still small to medium (0.44) in magnitude.

Effect sizes decreased in magnitude when looking at non-specific and specific treatments as comparator groups. For the non-specific treatment comparator group, effect sizes ranged from 0.18 (substance use frequency) to 0.42 (substance use quantity) for early follow-up. At late follow-up, the effect size for substance use frequency was 0.05 (there was no late follow-up substance use quantity reported for the non-specific comparator group). Thus, although smaller in overall magnitude compared to the minimal treatment group, CBT still had small to moderate effects at early-follow up when non-specific treatment was used as the comparator group. For the specific treatment comparator group, effect sizes were all close to zero regardless of follow-up period.

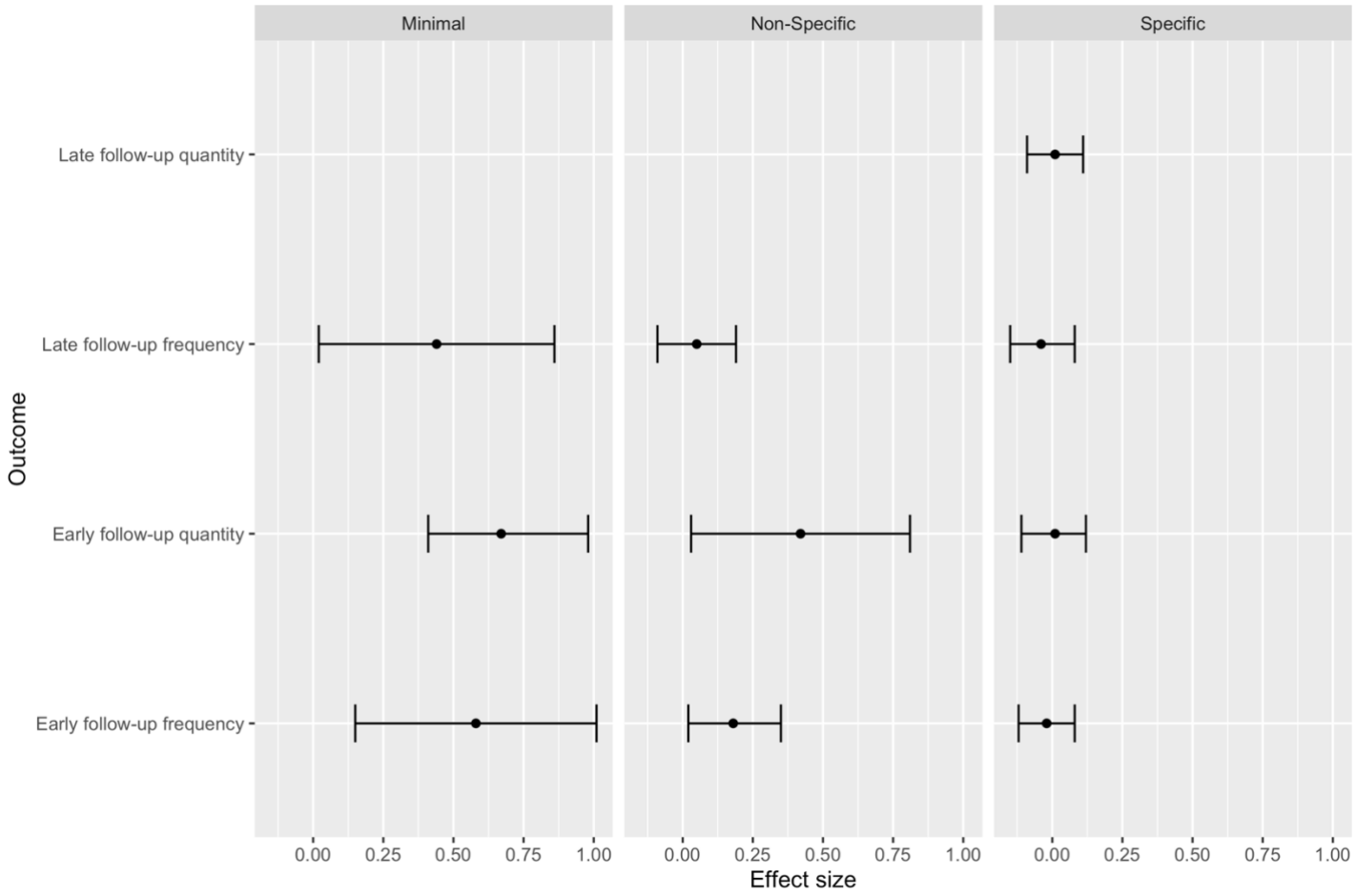

Figure 2. Effect size estimates (Hedge's *g*) with 95% confidence intervals for Magill et al., 2019 for each comparator group type.

For ease of comparison in the supplemental analyses (described later), we also averaged effect sizes between the non-specific and specific treatment comparator groups (to indicate active treatment effect sizes) and between substance use quantity and frequency (to indicate overall substance use effect sizes). These results are displayed in Figure 3. The general patterns previously observed are also true here. For example, all effect sizes are larger when the minimal (inactive) group is used as the comparator versus the active treatment group. Worth noting, however, is that there is a moderate effect of CBT on early follow-up quantity (0.22) when compared to active treatments.

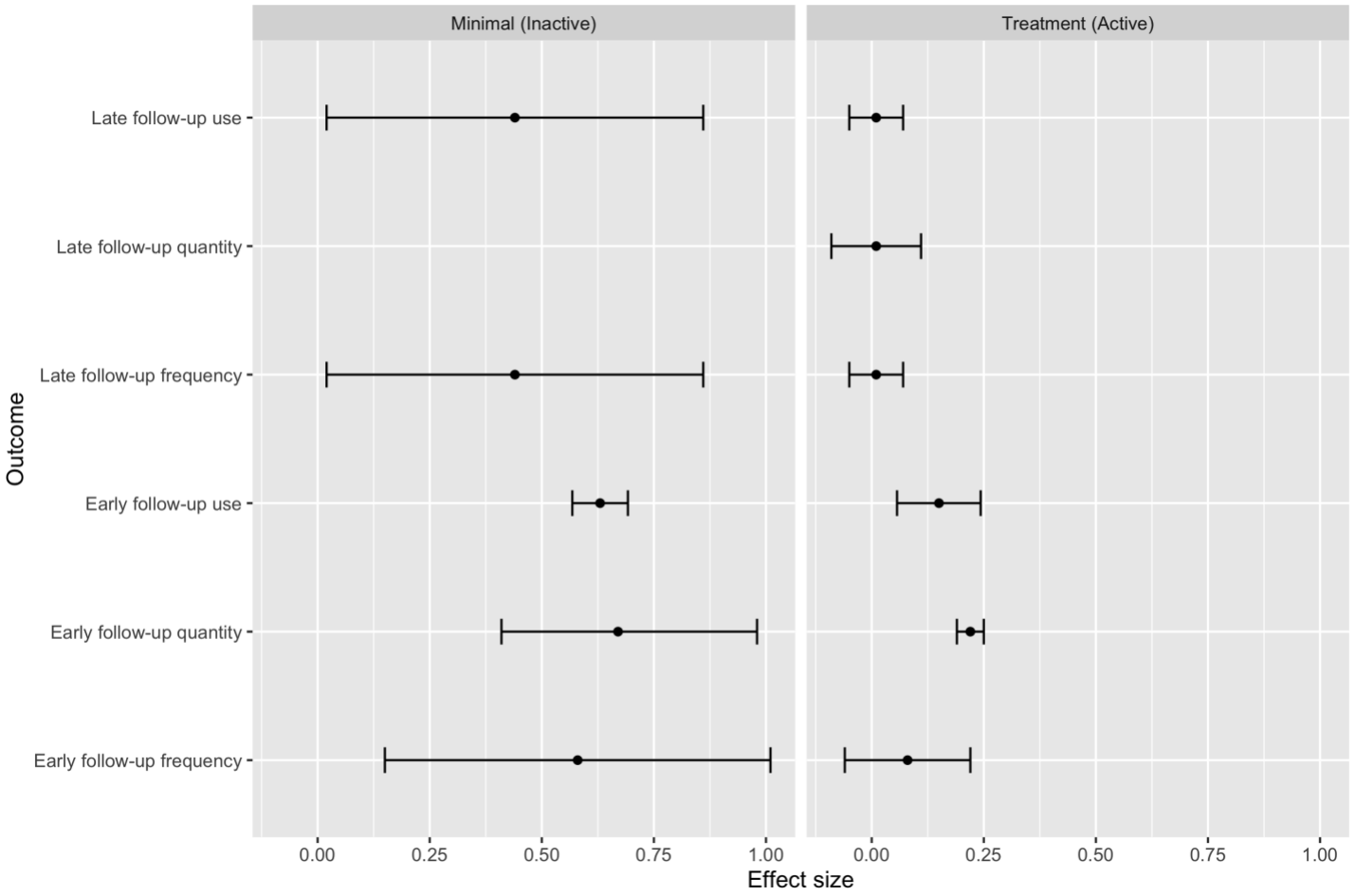

Figure 3. Effect size estimates (Hedge's *g*) with 95% confidence intervals for Magill et al., 2019 for minimal versus active treatment comparator groups. Treatment (Active) represents an aggregate of effect sizes for the original non-specific and specific treatment comparator groups. "Use" is the average of quantity and frequency, where available.

4.5 Sensitivity Analyses

As a sensitivity analysis, the four reviews of low or critically low quality were considered. Effect sizes were first converted to a consistent metric (Hedge's *g*) and then combined by taking the mean of the effect sizes reported for a given outcome. Because the goal of the Tolin Evaluation is not to estimate an aggregated effect of the treatment, we maintained the approach taken in the CBT-I evaluation (Boness et al., 2020) whereby we took an average of the effect sizes. This may also be useful for minimizing, to the extent possible, bias in the precision of the overall effect sizes due to non-independence. Although it is difficult to know the full extent of non-independence on an estimate's variance, we include Supplemental Table 2 as an index of the extent to which the meta-analyses considered eligible for the purposes of this evaluation included the same primary study in their effect estimates. We expect variance estimates to be minimally biased given most of the estimates reported below were derived from single meta-analyses.

We report combined effect sizes for four outcomes: 1) combined post-treatment and follow-up substance use, 2) post-treatment substance use, 3) early follow-up (1-6 months post-treatment) substance use, and 4) late follow-up (6+ months post-treatment) substance use.

For combined post-treatment and follow-up substance use, only effect sizes from Magill and Ray, 2009 were used. Although these outcomes were reported for Windsor et al., 2015 and Irvin et al., 1999, Windsor et al. (2015) included the pre-post comparison in their aggregate and Irvin et al., 1999 reported on this outcome in the *r* metric but did not report on total sample size for the effect so we could not convert *r* to Hedge's *g* for inclusion. Based on Magill and Ray, 2009 the combined post-treatment and follow-up effect for CBT on substance use was large in magnitude (g=0.80, 95% CI=0.45-1.14) when compared to an

inactive comparison group and very small (*g*=0.14, 95% CI=0.13-0.15) when compared to an active treatment comparison group.

The only study that reported on post-treatment substance use was Dutra et al., 2009. This study suggested a small to moderate effect of CBT on substance use at post-treatment compared to an inactive control comparison group (*g*=0.30, 95% CI = 0.27-0.33).

For early follow-up substance use, Irvin et al., 1999 and Magill et al., 2019 reported on this outcome. However, because Irvin et al., 1999 reported on this outcome in the *r* metric but did not report on total sample size for the effect, we could not convert *r* to Hedge's *g* for inclusion. Thus, the estimate for the effect of CBT on substance use at early follow-up is equivalent to what is reported in the main analyses above such that there is a small effect of CBT on substance use when compared to active treatment (*g*=0.15, 95% CI=0.06-0.24) and a moderate to large effect of CBT on substance use when compared to inactive, or minimal treatment (*g*=0.63, 95% CI=0.57-0.69).

For late follow-up substance use, Magill & Ray, 2009 and Magill et al., 2019 report on this outcome. It is worth noting that Magill & Ray, 2009 report effect sizes across all comparison group types and thus we could not separate these estimates into inactive versus active treatment comparison. Thus, comparison groups were combined across inactive (minimal) and active comparison groups for Magill et al., 2019. The overall effect of CBT on substance use at late follow-up across all comparison groups was small (*g*=0.17, 95% CI=0.08-0.25)

Together, these supplemental analyses support the main analyses such that CBT's effects on substance use are larger in magnitude when inactive (minimal) versus active treatment is the comparison group and these effects tend to diminish with time.

4.6 Judging the Quality of the Evidence for the Treatment

We designate the quality of the evidence for CBT in reducing substance use as being of moderate quality (see Table 3). This designation was made because we were only able to include one meta-analysis of adequate quality in the overall evaluation of the evidence (Magill et al., 2019) and this study was considered to have "moderate" quality per AMSTAR2 suggesting it had some limitations but not major flaws. However, we would also like to note that even those excluded based on quality produced similar results when included as part of sensitivity analyses. When considering the effect estimates extracted from Magill et al., 2019 confidence intervals were wide in some cases.

Table 3

Judging the Quality of the Evidence for CBT for Substance Use

Quality	Criteria
☐ *High quality*	All of the following: ● There is a wide range of studies included in the analyses with no major limitations. ● There is little variation between studies. ● The summary estimate has a narrow confidence interval.
✓ *Moderate quality*	At least one of the following: ● There are only a few studies, and some have limitations but not major flaws. ● There is some variation between studies, or the confidence interval of the summary estimate is wide.
☐ *Low quality*	Any of the following: ● The studies have major flaws. ● There is important variation between studies. ● The confidence interval of the summary estimate is very wide.

5. Consideration of Additional Contextual Factors

Reviewed meta-analyses lacked conclusive evidence across all relevant contextual factors. However, the reviewed meta-analyses taken together with additional literature provided support for four contextual factors: (1) CBT for substance use generates an effect that is similar to other well-studied treatments but has strong evidence for flexibility via technology-delivered CBT (e.g., Computer Based Training for CBT or CBT4CBT); (2) evidence supports the purported mechanism or active ingredient(s) of CBT for substance use; (3) CBT for substance use has been studied by a wide array of researchers (although with a mixed degree of allegiance to the treatment); (4) CBT for substance use demonstrates efficacy across several patient populations. Concerning the other two contextual factors, it appears that the effect sizes of CBT for substance use are comparable to other established and effective treatments, and there is inconclusive evidence that standard CBT for substance use has demonstrated good effects with marginalized populations. Overall, the four positive contextual factors support our overall treatment recommendation.

5.1 How does the treatment effect size compare to other established and effective treatments?

Based on the effect size estimates for Magill et al. (2019), the effect sizes for CBT on substance use outcomes are similar to active and specific behavioral interventions, such as Motivational Enhancement Therapy and Contingency Management. The effect sizes for CBT on substance use outcomes decreased over time within studies, with larger effect sizes observed at early follow-ups than late follow-ups. Other established interventions for substance use, including Motivational Enhancement Therapy and Contingency Management, demonstrate similar decreases in treatment effects over time (Benishek et al., 2014; Smedslund et al., 2011). There is some evidence that Contingency Management might result in superior long-term substance use outcomes (up to one-year post-treatment) over CBT; however, these meta-analytic findings were based on a limited number of studies, and results should be interpreted with caution (Ginley et al., 2021).

5.2 If the treatment effect size is similar to other established treatments, does the evaluated treatment differ in number of sessions, length, or cost?

As reviewed, the effect sizes for CBT on substance use outcomes are similar to active and specific interventions, such as Motivational Interviewing and Contingency Management (Magill et al., 2019). Although the number of sessions and treatment length of CBT for substance use varies widely, standard approaches are delivered over approximately 6-14 sessions, which is comparable to other established and efficacious treatments, such as Contingency Management and Twelve-Step Facilitation (Gates et al., 2016; Mattson et al., 1993). Conversely, Motivational Interviewing and Motivational Enhancement Therapy are typically delivered over fewer sessions (approximately 1-4 sessions) than CBT (Burke et al., 2003; Gates et al., 2016; Mattson et al., 1993). Few studies have examined the cost-effectiveness of CBT for substance

use, but limited extant findings indicate the cost-effectiveness of CBT for substance use (in combination with Motivational Enhancement Therapy) and technology-delivered CBT for substance is comparable to other established interventions for substance use, such as contingency management (Olmstead et al., 2010; Olmstead et al., 2007).

One additional considered contextual factor is the flexibility of the treatment modality. Although technology-delivered CBT was not included in the meta-analyses reviewed here, there is evidence that technology-delivered CBT for alcohol use as a standalone intervention demonstrates efficacy over minimal treatment (small effect size), and technology-delivered CBT for alcohol use as an adjunct to treatment as usual demonstrates efficacy over treatment as usual alone (small effect size) (Kiluk et al., 2019).[1] Technology-delivered CBT has also demonstrated efficacy for substances beyond alcohol, including cocaine use disorder and cannabis use disorder (Carroll, Ball, et al., 2008; Carroll, Kiluk, et al., 2014). The effect sizes observed for technology-delivered CBT on substance use outcomes are similar to effect sizes for all technology-delivered substance use treatments, including CBT and non-CBT interventions (e.g., Motivational Interviewing) (Rooke et al., 2010; Tait et al., 2013).

5.3 Is there evidence linking the treatment to the purported mechanism of change?

There is evidence that CBT for substance use exerts effects through hypothesized mechanisms of behavior change, including increased coping skills and self-efficacy. However, it is currently unclear the extent to which these mechanisms are unique to CBT versus common mechanisms of psychosocial treatments for substance use (Magill et al., 2020). Irvin et al. (1999) also found a medium effect (r=0.48) for RP (combining all follow-up timepoints and combining pre- to post-treatment effects and all comparator groups) on psychosocial adjustment outcomes. Notably, the definition of psychosocial adjustment outcomes included measures of purported mechanisms of change, such as self-efficacy, coping skills, and social and problem-solving skills. A recent meta-analysis of substance use interventions on emotional outcomes indicated that CBT was not statistically significantly associated with reductions in emotional distress, and that mindfulness-based and affect-regulation interventions had greater efficacy in reducing emotional distress than CBT (Kang, Fairbarn, & Ariss, 2019). However, a relatively small number of studies examining CBT (n = 4) were evaluated, as compared to the studies evaluating mindfulness-based (n = 11) and affect-regulation (n = 6) interventions. Overall, it's also worth acknowledging there is some evidence for the impact of CBT for substance use on psychosocial (or functional) outcomes, but this evidence is mixed and limited.

5.4 Is there evidence that supports treatment effectiveness in marginalized populations?

Windsor et al. (2015) compared the effect of CBT on substance use outcomes for studies with predominantly non-Hispanic White participants versus studies with predominantly Hispanic and/or Black samples. This meta-analysis concluded that effects sizes of CBT versus comparison groups were similar for studies enrolling predominantly non-Hispanic White and Hispanic and/or Black samples. However, although the effect sizes of CBT pre- to post-intervention were large and statistically significant in both groups, these effects were larger (indicating greater change) for studies enrolling predominantly non-Hispanic White versus Hispanic and/or Black samples. Windsor et al. (2015) also identified significant weaknesses in the literature, including a paucity of studies comparing retention and engagement rates by racial/ethnic identity, and few studies enrolling primarily Hispanic and/or Black samples. In addition, Magill et al., 2019, found that the percentage of participants who identified as white was not associated with early follow-up substance use frequency effect size (they only examined effect sizes with enough heterogeneity and a large enough sample size to examine subgroup effects). Overall, there is insufficient evidence that standard CBT for substance use has demonstrated good effects with marginalized populations. Further, there exist more general criticisms of the use of CBT as an approach among marginalized groups because it has been developed and tested in overwhelmingly White samples and overlooks cultural values that are likely to be held by marginalized groups (e.g., interdependence over personal independence; see Hays, 2009). Cultural adaptations of CBT, including modifications of setting, language, and content, to improve

[1] It is worth noting that Kiluk et al., 2019 was not captured by our original search strategy, conceivably because it focused on technology- or computer-delivered CBT rather than in-person CBT. Thus, we include Kiluk et al., 2019 in the consideration of contextual factors for the current evaluation but do not extract effect sizes for aggregation.

access, acceptability and efficacy in marginalized groups have shown initial promise, but will require additional research (e.g., Jordan et al., 2021; Paris et al., 2018).

5.5 Has the treatment been studied by a wide array of researchers without a strong allegiance to the treatment?

Previous studies assessing treatment allegiance have operationalized allegiance as one or more authors having developed this intervention and/or supervising or training the therapists delivering the study intervention (Dragioti et al., 2015). Based on studies included in the eligible meta-analyses (see Supplemental Table 2), CBT for substance use has been studied by several independent research groups with mixed allegiance to the treatment.

5.6 Is there evidence that supports treatment effectiveness across several patient populations?

Based on studies included in eligible quantitative reviews, CBT of substance use is generally efficacious across several patient populations, including those with varying primary substance use disorders (e.g., cannabis, alcohol, cocaine), those with co-occurring disorders (e.g., borderline personality disorder, posttraumatic stress disorder), those receiving adjunctive pharmacotherapy, and those in varying settings (e.g., community sample, specialty substance use or mental health clinic, medical setting, college setting, criminal justice setting, etc.) (Dutra et al., 2008; Magill et al., 2019). Several meta-analyses included in the present evaluation explicitly examined these patient factors as moderators of CBT effects or conducted subgroup analyses. Magill et al., 2019, concluded that primary substance was not associated with early follow-up substance use frequency effect size (they only examined effect size with enough heterogeneity and a large enough sample size to examine subgroup effects). Conversely, Irvin et al., 1999 found that RP was more efficacious for alcohol and polysubstance use than smoking or cocaine use, while Magill & Ray, 2009 found that effect sizes for CBT across substance type were similar (and small), except for cannabis, for which the effect size was moderate. Although the primary quantitative review we examined (Magill et al., 2019) excluded those with adjunctive pharmacotherapy, therefore indicating the efficacy of CBT alone, meta-analyses that included adjunctive interventions found that CBT + pharmacotherapy generally had greater effect sizes than CBT alone (Irvin et al., 1999; Magill & Ray, 2009). Other patient population factors that were examined as moderators or using a subgroup approach but were not significantly related to effect sizes included setting of treatment (Irvin et al., 1999), co-occurring psychological disorders (Magill & Ray, 2009), and treatment format (i.e., group vs. individual; Irvin et al., 1999, Magill et al., 2019).

Table 4

Additional contextual factors considered in increasing or decreasing the GRADE recommendation for CBT for Substance Use

Positive	Negative
☐ Treatment appears superior to other established and effective treatment(s)	☐ There are other psychological treatments that have well-documented and much larger effects
✓ The treatment generates an effect that is similar to other well-studied treatments and has strong evidence for flexibility via technology-delivered CBT	☐ The treatment generates an effect that is similar to other well-studied treatments, but requires a very large number of sessions or length of time to generate the same effect at a much higher cost
✓ Evidence supports the purported mechanism or active ingredient(s) of treatment	☐ Evidence fails to support the purported mechanism or active ingredient(s) of treatment
☐ Treatment has demonstrated good effects with marginalized groups	☐ Treatment has demonstrated weak effects with marginalized groups
☐ Treatment has been studied by a wide array of researchers without strong allegiance to the treatment	☐ Treatment has been studied by a narrow array of researchers with strong allegiance to the treatment
✓ Other: Demonstrated efficacy across several patient populations	☐ Other:

Note. This table identifies additional positive contextual factors supported by the CBT for substance use literature and was adapted from Tolin et al., 2015. Lack of identification of a positive or negative assessment of a contextual factor indicates that there is not enough data to make a firm conclusion in this category for CBT for substance use.

6. Overall Treatment Recommendation

Table 5

Overall Treatment Recommendation for CBT for Substance Use

Recommendation	Criteria
☐ *Very strong recommendation*	All of the following: ● There is high-quality evidence that the treatment produces a clinically meaningful effect on symptoms of the disorder being treated ● There is high-quality evidence that the treatment produces a clinically meaningful effect on functional outcomes ● There is high-quality evidence that the treatment produces a clinically meaningful effect on symptoms and/or functional outcomes at least three months after treatment discontinuation ● At least one well-conducted study has demonstrated effectiveness in non-research settings
✓ *Strong recommendation*	At least one of the following: ● There is moderate- to high-quality evidence that the treatment produces a clinically meaningful effect on symptoms of the disorder being treated ● There is moderate- to high-quality evidence that the treatment produces a clinically meaningful effect on functional outcomes
☐ *Weak recommendation*	Any of the following: ● There is only low- or very low-quality evidence that the treatment produces a clinically meaningful effect on symptoms of the disorder being treated ● There is only low- or very low-quality evidence that the treatment produces a clinically meaningful effect on as well as on functional outcomes ● There is moderate- to high-quality evidence that the effect of the treatment, although statistically significant, may not be of a magnitude that is clinically meaningful

Note. This table was adapted from Tolin et al., 2015.

Narrative Summary of GRADE Recommendation, Including Contextual Factors

There is moderate quality evidence that cognitive behavioral therapy for substance use produces small to moderate effects on substance use when compared to inactive treatment This remains true even when effect sizes from studies considered to have low or critically low quality are considered. There is also some evidence, although from a dated study that is considered to have “critically low” quality per AMSTAR2, that CBT for substance use may have an effect on psychosocial/functional outcomes. As such, based on the criteria outline by Tolin and colleagues (2015), the current status of the literature merits a **strong** recommendation of CBT for substance use.

Given the paucity of data on related psychosocial outcomes, we cannot conclude that the treatment produces a clinically meaningful effect on functional outcomes at this time.

Our **strong** recommendation is further strengthened by several contextual factors:

1. The treatment generates an effect that is similar to other well-studied treatments for substance use and substance use disorders. CBT for substance use can be flexibly delivered via computerized approaches.
2. The treatment exerts effects through hypothesized mechanisms of behavior change, including increased coping skills and self-efficacy although it is currently unclear the extent to which these mechanisms are unique to CBT versus common mechanisms of psychosocial treatments for substance use.
3. There is evidence of efficacy across various patient populations. However, more meta-analytic research on the efficacy and acceptability of CBT for substance use with marginalized populations is required.

References

* = meta-analysis included in current evaluation

Benishek, L. A., Dugosh, K. L., Kirby, K. C., Matejkowski, J., Clements, N. T., Seymour, B. L., & Festinger, D. S. (2014). Prize-based contingency management for the treatment of substance abusers: A meta-analysis. *Addiction*, *109*(9), 1426-1436.

Burke, B. L., Arkowitz, H., & Menchola, M. (2003). The efficacy of motivational interviewing: a meta-analysis of controlled clinical trials. Journal of consulting and clinical psychology, *71*(5), 843.

Carroll, K. M., Ball, S. A., Martino, S., Nich, C., Babuscio, T. A., Nuro, K. F., ... & Rounsaville, B. J. (2008). Computer-assisted delivery of cognitive-behavioral therapy for addiction: a randomized trial of CBT4CBT. *American Journal of Psychiatry*, *165*(7), 881-888.

Carroll, K. M., Kiluk, B. D., Nich, C., Gordon, M. A., Portnoy, G. A., Marino, D. R., & Ball, S. A. (2014). Computer-assisted delivery of cognitive-behavioral therapy: efficacy and durability of CBT4CBT among cocaine-dependent individuals maintained on methadone. *American journal of Psychiatry*, *171*(4), 436-444.

Dragioti, E., Dimoliatis, I., Fountoulakis, K. N., & Evangelou, E. (2015). A systematic appraisal of allegiance effect in randomized controlled trials of psychotherapy. *Annals of General Psychiatry, 14*(1), 1-9.

*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 Psychiatry*, *165,* 179-187.

Gates, P. J., Sabioni, P., Copeland, J., Le Foll, B., & Gowing, L. (2016). Psychosocial interventions for cannabis use disorder. *Cochrane Database of Systematic Reviews*, (5).

Ginley, M. K., Pfund, R. A., Rash, C. J., & Zajac, K. (2021). Long-term efficacy of contingency management treatment based on objective indicators of abstinence from illicit substance use up to 1 year following treatment: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *89*(1), 58.

Hays, P. A. (2009). Integrating Evidence-Based Practice, Cognitive-Behavior Therapy, and Multicultural Therapy: Ten Steps for Culturally Competent Practice. *Professional Psychology: Research and Practice*, *40*(4), 354–360.

*Irvin, J. E., Bowers, C. A., Dunn, M. E., & Wang, M. C. (1999). Efficacy of relapse prevention: a meta-analytic review. *Journal of Consulting and Clinical Psychology*, *67*(4), 563-570.

Jordan, A., Babuscio, T., Nich, C., & Carroll, K. M. (2021). A feasibility study providing substance use treatment in the Black church. *Journal of Substance Abuse Treatment*, *124*, 108218.

Kang, D., Fairbairn, C. E., & Ariss, T. A. (2019). A meta-analysis of the effect of substance use interventions on emotion outcomes. *Journal of Consulting and Clinical Psychology, 87*(12), 1106.

Kiluk, B. D., Ray, L. A., Walthers, J., Bernstein, M., Tonigan, J. S., & Magill, M. (2019). Technology-delivered cognitive-behavioral interventions for alcohol use: a meta-analysis. *Alcoholism: clinical and experimental research*, *43*(11), 2285-2295.

*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,* 516-527.

*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 & Clinical Psychology*, *87*(12), 1093–1105.

Magill, M., Tonigan, J. S., Kiluk, B., Ray, L., Walthers, J., & Carroll, K. (2020). The search for mechanisms of cognitive behavioral therapy for alcohol or other drug use disorders: A systematic review. *Behaviour Research and Therapy*, *131*, 103648.

Mattson, M. E., Allen, J. P., Miller, W. R., Hester, R. K., Connors, G. J., Rychtarik, R. G., ... & DelBoca, F. K. (1993). Project MATCH: Rationale and methods for a multisite clinical trial matching patients to alcoholism treatment. *Alcoholism: Clinical and Experimental Research*, *17*(6), 1130-1145.

McHugh, R. K., Hearon, B. A., & Otto, M. W. (2010). Cognitive behavioral therapy for substance use disorders. *Psychiatric Clinics*, *33*(3), 511-525.

Olmstead, T. A., Ostrow, C. D., & Carroll, K. M. (2010). Cost-effectiveness of computer-assisted training in cognitive-behavioral therapy as an adjunct to standard care for addiction. *Drug and alcohol dependence*, *110*(3), 200-207.

Olmstead, T. A., Sindelar, J. L., Easton, C. J., & Carroll, K. M. (2007). The cost-effectiveness of four treatments for marijuana dependence. *Addiction*, *102*(9), 1443-1453.

Paris, M., Silva, M., Añez-Nava, L., Jaramillo, Y., Kiluk, B. D., Gordon, M. A., ... & Carroll, K. M. (2018). Culturally adapted, web-based cognitive behavioral therapy for Spanish-speaking individuals with substance use disorders: A randomized clinical trial. *American Journal of Public Health*, *108*(11), 1535-1542.

Rooke, S., Thorsteinsson, E., Karpin, A., Copeland, J., & Allsop, D. (2010). Computer-delivered interventions for alcohol and tobacco use: A meta-analysis. *Addiction*, *105*(8), 1381-1390.

Smedslund, G., Berg, R. C., Hammerstrøm, K. T., Steiro, A., Leiknes, K. A., Dahl, H. M., & Karlsen, K. (2011). Motivational interviewing for substance abuse. *Campbell Systematic Reviews*, *7*(1), 1-126.

Tait, R. J., Spijkerman, R., & Riper, H. (2013). Internet and computer based interventions for cannabis use: a meta-analysis. *Drug and Alcohol Dependence*, *133*(2), 295-304.

*Windsor, L. C., Jemal, A., & Alessi, E. J. (2015). Cognitive behavioral therapy: A meta-analysis of race and substance use outcomes. *Cultural Diversity and Ethnic Minority Psychology*, *21*(2), 300–313.

Supplemental Table 1. *Raw Effect Size Estimates.*

Outcome	Comparator Group	Effect Size	95% Confidence Interval (lower)	95% Confidence Interval (upper)
Late follow-up frequency	Minimal	0.44	0.02	0.86
Early follow-up quantity	Minimal	0.67	0.41	0.98
Early follow-up frequency	Minimal	0.58	0.15	1.01
Late follow-up frequency	Non-Specific	0.05	-0.09	0.19
Early follow-up quantity	Non-Specific	0.42	0.03	0.81
Early follow-up frequency	Non-Specific	0.18	0.02	0.35
Late follow-up quantity	Specific	0.01	-0.09	0.11
Late follow-up frequency	Specific	-0.04	-0.15	0.08
Early follow-up quantity	Specific	0.01	-0.11	0.12
Early follow-up frequency	Specific	-0.02	-0.12	0.08

Note. These estimates correspond to values depicted in Figure 2.

Supplemental Table 2. *Overlap Among Primary Studies Included in Current Evaluation.*

Primary Study	Meta-analysis					*N* of meta-analyses including each	% Meta-analyses included
	Dutra et al., 2008	Irvin et al., 1999	Magill & Ray, 2009	Magill et al., 2019	Windsor et al., 2015		
Abbott et al., 1998	X					1	20
Annis & Peachey, 1992		X	X			2	40
Anton et al., 1999					X	1	20
Anton et al., 2005			X		X	2	40
Anton et al., 2006			X			1	20
Ashkanazi, 1990		X				1	20
Azrin et al., 1994	X					1	20
Babor, 2004			X			1	20
Balldin et al., 2003			X			1	20
Bennett et al., 2005			X			1	20
Bickel et al., 1997	X					1	20
Bowen et al., 2014				X		1	20
Brown et al., 2002			X	X		2	40
Brown et al., 2006					X	1	20
Budney et al., 2000	X		X			2	40
Budney et al., 2006			X	X	X	3	60
Burtscheidt et al., 2002			X	X		2	40
Carroll et al., 1991			X	X		2	40
Carroll et al., 1994	X	X	X			3	60
Carroll et al., 1998	X		X			2	40
Carroll et al., 2000			X			1	20
Carroll et al., 2001	X					1	20
Carroll et al., 2002	X					1	20
Carroll et al., 2004			X			1	20
Carroll et al., 2012					X	1	20
Carroll, 1988		X				1	20
Chaney et al., 1978		X				1	20
Chutuape et al., 2001	X					1	20
Conrod et al., 2000			X			1	20
Cooney, 1991			X			1	20
Copeland et al., 2001	X		X			2	40
Crits-Christoph et al., 1999	X		X			2	40
Dawe et al., 2002				X		1	20
Donovan et al., 1988			X	X		2	40
Downey et al., 2000	X					1	20
Epstein et al., 2003			X		X	2	40
Gilbert et al., 2006			X			1	20
Goldstein et al., 1989		X				1	20

Gottheil et al., 2002	X					1	20
Gregory, 1984		X				1	20
Hall et al., 1984		X				1	20
Hammarberg et al., 2004			X			1	20
Hawkins et al., 1986		X	X			2	40
Hawkins et al., 1989		X	X			2	40
Heather et al., 2000				X		1	20
Heinala et al., 2001			X			1	20
Hien et al., 2004	X					1	20
Higgins et al., 1993	X					1	20
Hill et al., 1993		X				1	20
Iguchi et al., 1997	X					1	20
Ito et al., 1988		X				1	20
Jaffee et al., 1996			X			1	20
Jones et al., 1982			X	X		2	40
Jones et al., 2004	X					1	20
Kadden et al., 1989			X	X		2	40
Kadden et al., 2001			X	X		2	40
Katz et al., 2002	X					1	20
Kelly et al., 2000			X			1	20
Kennedy et al., 2012					X	1	20
Kivlahan et al., 1990				X		1	20
Knight et al., 1994		X				1	20
Kosten et al., 2003	X					1	20
Kranzler et al., 1995		X				1	20
Lanza et al., 2014				X		1	20
Linehan et al., 1999	X					1	20
Linehan et al., 2002	X					1	20
Litt et al., 2016				X		1	20
Litt, 2003			X			1	20
Lydecker et al., 2010					X	1	20
Maisto et al., 1995		X				1	20
Maude-Griffin et al., 1998			X	X		2	40
McAuliffe, 1990	X		X	X		3	60
McKay et al., 1997			X	X		2	40
McKay et al., 2004				X		1	20
McKay et al., 2010				X		1	20
Messina et al., 2003			X			1	20
Monti et al., 1990			X			1	20
Monti et al., 1993			X			1	20
Monti et al., 1997			X	X		2	40
Monti et al., 2001			X			1	20

Morgenstern et al., 2001			X	X	X	3	60
O'Connell, 1987		X				1	20
O'Farrell et al., 1993		X				1	20
O'Malley et al., 1992		X	X			2	40
Oslin et al., 2008					X	1	20
Papas et al., 2011				X		1	20
Peters et al., 1993		X				1	20
Petry & Martin, 2004	X					1	20
Pollack et al., 2002	X		X			2	40
Project MATCH Research Group, 1997			X	X	X	3	60
Rawson et al., 2001	X					1	20
Rawson et al., 2002			X			1	20
Rawson et al., 2006			X			1	20
Roffman et al., 1988	X					1	20
Rohsenow et al., 1991			X			1	20
Rohsenow et al., 2000			X			1	20
Rohsenow et al., 2001			X			1	20
Rohsenow et al., 2004			X			1	20
Rosenblum et al., 2005			X			1	20
Rowan-Szal et al., 2005			X		X	2	40
Sandahl & Ronnberg, 1990		X				1	20
Sandahl et al., 2004			X	X		2	40
Schmitz et al., 2001			X			1	20
Schmitz et al., 2002a	X					1	20
Schmitz et al., 2004			X			1	20
Schmitz et al., 2009					X	1	20
Shakeshaft et al., 2002				X		1	20
Shoptaw et al., 2005					X	1	20
Sigmon et al., 2004	X					1	20
Silverman et al., 1996	X					1	20
Silverman et al., 2004	X					1	20
Sinha et al., 2003	X					1	20

Sitharthan et al., 1997				X		1	20
Smout et al., 2010				X		1	20
Sobell et al., 1995		X	X			2	40
Stephens et al., 1994			X	X		2	40
Stephens et al., 2000	X		X	X	X	4	80
Stevens & Hollis, 1989		X				1	20
Stevens et al., 1993		X				1	20
Stitzer et al., 1992	X					1	20
Supnick & Colletti, 1984		X				1	20
Thorton et al., 2003				X		1	20
Tucker et al., 2004			X			1	20
Wells et al., 1994		X				1	20
Wetzel et al., 2004			X			1	20
Zelman et al., 1992		X				1	20

Note. Double-coded by CB and VV.