

A Closer Look at Exclusion Criteria in Randomized Controlled  
Psychotherapy Trials for Adult Depression

Anneliese M. Schenk

Ohio State University

Author Note

Anneliese M. Schenk, Department of Psychology, Ohio State University.

I thank Daniel Strunk who advised and guided this thesis and August Gonzalez for assistance in the data collection process.

Address correspondence of the results to Anneliese Schenk, Department of Psychology, Ohio State University, 1835 Neil Avenue, Columbus OH, 43210. [schenk.67@osu.edu](mailto:schenk.67@osu.edu)

### Abstract

A common concern regarding clinical trials evaluating treatments for depression is findings may not generalize to the population of interest due to the use of restrictive exclusion criteria. To evaluate the exclusion criteria used in such research, we examined a large body of reports of randomized controlled trials of psychotherapies for adult depression. We analyzed the rates of four major types of exclusions: suicide related exclusions, an exclusion for when depression is not deemed the primary or principal diagnosis, exclusions due to comorbid diagnoses, and exclusions related to substance use. Data were drawn from a compilation of randomized controlled and comparative trials of psychotherapy on adult depression assembled by Professor Pim Cuijpers. Our analyses include 332 of the original 358 studies, as we excluded prevention studies, meta analyses, reports not available in English, and those that we could not access. Two raters coded each of these studies. Reliability for the raters' judgements ranged from moderate to high (Cohen's kappa of .52 to 1.00). Of the studies examined, 42% used a suicide related exclusion, 72% excluded for the presence of specific comorbid disorders, 63% excluded participants who did not have depression as a primary / principal diagnosis, and 52% excluded participants on the basis of substance use. Within those broader categories, the following specific exclusions were found to be used increasingly over time: use of a measure for operationalizing a suicide-related exclusion, self-harm, any bipolar spectrum disorder, current alcohol abuse, and current drug abuse. In contrast, four exclusions were found to be used with reduced frequency over time: requiring hospitalization, antisocial personality disorder (ASPD), having no disorder-based exclusions, and past alcohol abuse. These changes show that exclusion criteria have changed over the course of the last fifty years. To our knowledge, these changes have been unidentified to date. While researchers may disagree about what exclusion criteria should be

used in trials, specific differences have gotten little attention in the literature, perhaps preventing a consensus about under what circumstances various criteria should be used. We hope our work can help to prompt greater consideration of the use of specific exclusion criteria.

## A Closer Look at Exclusion Criteria in Randomized Controlled Psychotherapy Trials for Adult Depression

It is widely recognized that when designing an experiment there are often compromises between external and internal validity, and clinical trial investigators have generally tended to favor the latter. High levels of internal validity mean that the experiment properly manipulated its variables and allows for valid causal inferences, while high levels of external validity mean that the results of the study are likely to generalize to the population of interest. A number of researchers have suggested that the goal of achieving higher internal validity has encouraged researchers to use overly restrictive inclusion / exclusion criteria (see Westen, Novotny, & Thompson-Brenner, 2004). Often researchers “associate restrictive exclusion / inclusion criteria with lower external validity and higher internal validity” (Lorenzo-Luaces, Zimmerman, & Cuijpers, 2018), causing them to exclude participants that may belong to the population of interest in order to avoid confounding variables.

For a disorder such as depression that is common, heterogenous and often comorbid with other conditions, this view appears especially relevant. Depression is commonly comorbid with other disorders. For example, Brière, Rohde, Seeley, Klein, and Lewinsohn (2014) found that 40% of participants with a lifetime history of major depressive disorder (MDD) had a history of alcohol use disorder (AUD) (and that 59% of those with a lifetime history of AUD had a history of MDD). Additionally, 27% of the individuals in the study with concurrent MDD and AUD were found to have had MDD first, perhaps suggesting that MDD was primary in this subset of cases (Brière, Rohde, Seeley, Klein, & Lewinsohn, 2014). Likewise, surveys have found that 6.1% of individuals with MDD also experience drug abuse or dependence in their lifetime, and 1.5% have in the last year (Grant, 1995).

Concerns surrounding the above findings have led some to question the external validity of randomized controlled trials (RCTs; see Westen, Novotny, & Thompson-Brenner, 2004) and to call for a shift away from treatment packages towards providing clinicians with individual empirically supported principles that they could integrate into treatments. Discussions of the limitations of exclusion criteria would benefit from a more precise characterization of what criteria have been and are being used in trials. In this thesis, we will examine exclusion criteria trends over a fifty-year span of clinical trials, including both RCTs (trials with a control group) and comparative efficacy studies (trials lacking a control group).

Though restrictive exclusion criteria may be problematic for the generalizability of research on the population of those with depression, its exact effects are difficult to discern. Other research such as that by Lorenzo-Luaces, Zimmerman, and Cuijpers (2018) has shown that though there are issues with the inclusion / exclusion criteria in MDD research, these concerns might be greatest for trials of antidepressant medication. Lorenzo-Luaces, Zimmerman, and Cuijpers (2018) analyzed the differences in the use of inclusion / exclusion criteria between psychotherapy and antidepressant studies. They coded sixteen studies for their participant exclusion / inclusion criteria and found that psychotherapy studies tend to be more inclusive than antidepressant medication studies overall. Compared to psychotherapy studies, antidepressant medication trials tended to more often exclude for brief episodes, low symptom severity, and specific diagnoses other than depression. Both types of studies excluded for psychotic symptoms, substance abuse / dependence, and suicidality at equally high rates, with the most common exclusion for both being milder depressive symptoms. Insofar as problems related to inclusion / exclusion criteria differ across different treatment modalities, the quality of inferences from meta-analyses comparing across these studies is likely to be compromised.

Wiltsey, DeRubeis, Christoph, and Brody (2003) suggested that a high rate of exclusion is not in and of itself problematic for the generalizability of the clinical trial literature. They point out that patients who have constellations of problems that lead them to be excluded from one trial may still be eligible for another trial. Concerns about generalizability of the clinical trial literature are warranted if there are patients who have a clinical presentation that is unrepresented in the clinical trial literature. If so, such patients are understudied and results of clinical trials are not available to inform their care. In order to determine if individuals who were excluded from a clinical trial would be deemed ineligible from all clinical trials conducted at the time, Wiltsey and colleagues had raters evaluate 347 patients' clinical charts from a managed care sample and rate on a 4-point Likert scale if each patient would qualify for an RCT for any of 10 psychological disorders. Of patients with a diagnosis studied in the literature, 80% would have been eligible for one or more trials. However, 58% of the sample was composed of patients with diagnoses for which no trials existed, such as adjustment disorder. When a patient had a diagnosis that had been studied but was not eligible for any trial, they were often excluded on the basis of having low severity or substance abuse or dependence. This work suggests that a pattern of high exclusion rates from trials may not provide the information needed to assess the generalizability of the literature. Moreover, the specific findings suggest more specific areas to focus on to improve the extent to which the literature can more fully inform care for a wide variety of clients.

Wiltsey, DeRubeis, Christoph, and Rothman built on their previous work in a follow-up study (2005) exploring the exclusion rates of RCTs specifically in MDD trials. They pulled records for potential participants excluded from MDD studies at the University of Pennsylvania between 1995 and 2001. They had graduate students analyze if the formerly excluded

participants would be eligible for any published studies evaluating treatments for MDD, panic disorder, generalized anxiety disorder (GAD), post-traumatic stress disorder (PTSD), alcohol dependence, substance dependence, bulimia, obsessive-compulsive disorder (OCD), social phobia, or schizophrenia. They found that among those excluded from previous studies participants with substance use disorders and social anxiety disorder were at increased risk for exclusion in all RCTs. Such analyses help to address the question of whether there are patients who are excluded not only from one trial, but have clinical problems that are excluded from all trials. Of course, all of this work is dependent upon clearly defined and well-measured inclusion and exclusion criteria. Whether researchers are clear about their criteria and how they are assessing these criteria is important to understand for that reason.

Both the work of Lorenzo-Luaces and colleagues as well as Wiltsey and colleagues focused on carefully examining relatively small numbers of studies, and each acknowledged that broader analysis of the larger literature is needed. In light of this prior work, and in order to answer the questions surrounding how researchers assess inclusion / exclusion criteria and how the use of exclusion criteria may have changed over time within depression research, we sought to examine inclusion and exclusion criteria in a large collection of randomized trials. Through this research we will analyze the use of inclusion / exclusion criteria in trials of psychotherapies for adults with depression and how use of these criteria may have changed over time.

We will examine four distinct areas of exclusion criteria within adult depression RCTs: suicide related exclusions, disorder exclusions, exclusions related to depression not being primary, and substance use exclusions. Additionally, we will evaluate the possibility that use of these criteria has changed over time. We expect to find a high rate of studies using substance use related exclusions. We also expect to see a decline in recent years based on previous researchers'

recommendations that the literature be more inclusive of these participants. Similarly, we expect to find high rates of axis I disorder related exclusions with a decrease in their usage in more recent years. We expect to see a moderate use of exclusions related to depression not being primary and suicide related exclusions. We were also interested in exploring the extent to which investigators use clear measurement-based criteria for assessing these exclusions.

## **Method**

### **Identification and Selection of Studies**

The data set used for this study comes from a compilation of randomized controlled and comparative trials of psychotherapy on adult depression created by Professor Pim Cuijpers from the University of Amsterdam. In total, the data set consists of 358 studies. The data set, and the criteria used in its creation can be accessed at <http://www.evidencebasedpsychotherapies.org/>. The studies included provide a list of all studies on adult depression conducted from 1975 through January 2012 that were found using literature searches in the databases PubMed, Embase, PsycINFO, and the Cochrane Register of Controlled Trials. From this initial search Cuijpers excluded duplicates, trials on maintenance, collaborative care, and those that were not published in English, German, Spanish, or Dutch. Cuijpers and his team independently coded the following for each study: specific psychotherapy or pharmacotherapy provided, the control and experimental group(s) condition(s), the number of participants in each trial, and effect sizes associated with each of the studies.

Our only exclusion criteria for studies beyond those laid out by Cuijpers was that they must be available in English, must be accessible, and could not be either a prevention study or a meta-analysis. Ultimately, eleven studies were excluded due to not being available in English,

eleven for being prevention studies, two for being meta-analyses, and two for being inaccessible. This resulted in the final set of 332 studies.

### **Study Coding**

Two raters independently analyzed each study's method section to assess its inclusion and exclusion criteria. Each study was coded for what were expected to be common exclusion and inclusion criteria along with additional inquiries about how some criteria were assessed. Less common exclusion / inclusion criteria were coded by selecting an 'other' option which would then allow for an open-ended response. To determine inter-rater reliability Cohen's kappa scores were calculated for each individual variable within the exclusion criteria categories.

We coded for use of the following exclusion criteria in each study: suicide related exclusions, depression not being the primary / principal diagnosis, exclusions due to comorbid diagnoses other than substance use disorders, and exclusions related to substance abuse. Other criteria not analyzed for this report were also coded. These included: requirement of a medical or psychiatric condition other than depression, any participant demographic requirements, journal of publication, date of publication, continent of experiment, and additional exclusions related to medical problems. If data were available, we also recorded what diagnostic tools, if any, were used to assess participants' conditions.

### **Suicide Related Exclusions**

The following options were available to the coders when answering if a study contained any suicide related exclusions: requiring a current / previous ideation, imminent suicide risk (without measure specified), require hospitalization, suicidal ideation (without measure specified), past history of suicidal ideation exclusion, self-harm, score above a threshold on a

suicide measure, plan, and not specified. Coders were allowed to select multiple options to characterize the use of multiple exclusion criteria.

Use of suicide measures were coded to collect information on any measures used to assess suicidal risk as part of inclusion or exclusion criteria. In order for a study to be coded as including a suicide measure, its methods section had to mention that participants needed a specific score on a questionnaire to determine if they would be included in the study. Examples include the use of the Patient Health Questionnaire item 9, the suicide item on the Beck Depression Inventory II, and the Montgomery-Asberg Depression Rating Scale item 10.

A study was coded as having an exclusion for the presence of suicidal ideation (without measure specified) if within the methods section it mentioned participants being unable to have a suicidal ideation at the time of the study, but did not offer an operational definition of suicidal ideation. The same is true for the variable suicide risk (without measure specified) in which the phrase suicide risk appeared in the methods section, but the phrase was never defined or operationalized.

### **Comorbid Disorder Exclusions**

An exclusion criterion was coded within the category of Comorbid Disorder Exclusions when a study's methods section specified that individuals with a specific psychological disorder (other than disorders of substance use or abuse or eating disorders) would be excluded from participating in the study. Exclusions for physical ailments (e.g., cancer), neurocognitive disorders and organic mental disorders (e.g., dementia, Parkinson's Disease), and eating disorders were coded separately and are not analyzed within this paper. When coding for disorder exclusions the coders were allowed to select multiple of the following options: ADD/ADHD, ASPD, any personality disorder, cluster A personality disorders, any form of

bipolar disorder, borderline personality disorder (BPD), axis I disorders, psychosis, schizotypal, schizoaffective, and none. Additional categories were created as needed over the course of the rating process.

### **Depression Not Primary Exclusion**

Use of an exclusion criterion for depression not being the primary diagnosis was also coded. A diagnosis of depression was defined as either the study specifying that it required participants to have a diagnosis of depression or that a participant needed to meet research diagnostic criteria or criteria from the diagnostic and statistical manual of mental disorders.

### **Substance Use Exclusions**

A substance use exclusion was coded when studies excluded for conditions defined by the consumption of alcohol or drugs. It was possible for a study to be coded as any of the following seven substance exclusions; none, current alcohol abuse, current drug abuse, past alcohol abuse, past drug abuse, alcohol abuse timeframe not specified, and drug abuse timeframe not specified. If a study stated that a participant was excluded due to substance dependence or abuse, the study was coded as both current alcohol and current drug abuse. Additionally, studies that required participants to have substance dependence were coded as none, as this was an inclusion rather than an exclusion criterion.

## **Results**

### **Cohen's Kappa**

Given the categorical nature of the ratings, Cohen's kappa scores were calculated for each exclusion criteria category to assess inter-rater reliability. As shown in Table 1, the scores ranged from .66 to 1.00, indicating a moderate to strong level of inter-rater reliability.

### **Descriptive Statistics**

For suicide related exclusions, of the 332 total studies analyzed, 180 (54%) studies did not have any suicide related exclusion criteria. Three (1%) required participants to either have previously or currently have suicidal ideation in order to participate. Among the remaining 139 studies with a suicide related exclusion, the most common suicide related exclusion was suicidal ideation with 88 studies (63%), followed by 56 studies that excluded for suicide risk (40%), 22 that used a suicide measure based exclusion (16%), 14 that excluded due to having a suicide plan (10%), 10 that excluded for requiring hospitalization (7%), 3 that had a self-harm related exclusion (2%), and 18 that had history of suicide attempts and/or past ideation as an exclusion (13%). There was overlap among studies' exclusion criteria with 44 studies using two exclusion criteria, and 7 using three or more.

For disorder exclusions, of the 332 total studies analyzed, 99 (30%) did not exclude participants due to having another psychological disorder other than depression (this included psychological disorders other than substance use disorder or eating disorders). The remaining 233 (70%) studies did have at least one such exclusion. The exclusions were: 163 studies excluded for experiencing or having a history of psychosis (49%), 135 studies excluded for any bipolar spectrum disorder (41%), 46 studies excluded for panic or other anxiety disorders (14%), 39 studies excluded for ASPD (12%), 25 studies excluded for BPD (8%), 17 studies excluded for schizotypal personality disorder (5%), and 11 studies excluded for any personality disorder (3%).

For depression as not primary exclusions, of the 332 total studies, 208 (63%) studies excluded participants for not having a primary diagnosis of MDD, with the 124 remaining studies (37%) allowing participants to have a primary diagnosis other than MDD, typically by not having criteria related to what diagnosis is primary. The 208 studies that excluded

participants for not having a primary diagnosis of MDD were broken down into four categories. The most common category was studies requiring participants to have a non-primary diagnosis of MDD (at 126, 38%), followed by studies listing multiple exclusions related to a lack of diagnosis or MDD not being primary/principal (at 40, 12%), directly stating that depression is not primary / principal (at 31, 9%), and exclusions relative to a specific diagnosis (e.g., only excluded participants if a specific disorder was primary over MDD)(at 30, 9%).

Regarding substance use exclusions, 173 out of 332 total studies (52%) excluded participants for some form of substance abuse or dependence. These 173 studies were further broken down into the following categories: current alcohol abuse (105 studies, 60% of substance use exclusions), current drug abuse (109 studies, 63% of substance use exclusions), past alcohol abuse (15 studies, 9% of all substance use exclusions), past drug abuse (18 studies, 10% of all substance use exclusions), alcohol abuse timeframe not specified (34 studies, 20% of all substance use exclusions), and drug abuse timeframe not specified (39 studies, 23% of all substance use exclusions).

### **Frequency of Years**

We also ran Wilcoxon rank sum tests comparing the years of publication for studies with vs. without each exclusion criterion. This test allows us to determine if there is a difference in date of publication between studies that did vs. did not use exclusion criteria (see Table 2). In total we ran 29 Wilcoxon rank sum tests and found that 9 were statistically significant at the  $p < .05$  level. These differences suggest a relation between studies dates of publication and the exclusion criteria used.

As shown in Table 2, use of self-harm, excluding based on a suicide related measure, and requiring hospitalization all exhibited significant change over time. The exclusion for requiring

hospitalization tended to be used more often in earlier studies with the median year for studies using this exclusion being 1996 as compared the median year for studies not using the exclusion being 2001. For self-harm and suicide measures, the use of these exclusions was more common in recent studies. For self-harm, the median year of studies using the exclusion was 2010 as compared with the median year for studies not using the exclusion of 2001. Similarly, for suicide measures, the median year of studies using the exclusion was 2005 as compared with the median year for studies not using the exclusion of 2001.

Of the 11 disorder related exclusion variables examined, three showed significant changes in frequency over time: none (indicating no disorder-based exclusions), bipolar, and ASPD. Both the variable none and ASPD were used significantly more often in earlier studies. For none, the median year of studies with the exclusion (i.e., studies that did not exclude based on having a disorder other than MDD) was 1998 whereas the median year for the remaining studies was 2002. For ASPD, the median year of studies using this exclusion was 1999 and the median year for the remaining studies was 2001. Use of the bipolar exclusion was more common in recent studies. The median year of studies using a bipolar exclusion was 2002, whereas the median year for the remaining studies was 2000.

Of the seven substance use disorder variables examined, three showed significant changes over time: current alcohol abuse, current drug abuse, and past alcohol abuse. Past alcohol abuse was excluded more often in earlier studies, with the median year being 1995 for studies using this exclusion vs. the median year of 2001 for those not using the exclusion. Both current alcohol abuse and current drug abuse were used more often in later years. For both of these exclusions, the median year for studies using the exclusions was 2003, whereas the median for studies not using the exclusion being 2000.

### Discussion

One key finding from this study is that only 16% of all studies with suicide related exclusions clearly defined their methods for that exclusion and were coded as using suicide measures. The remaining 84% of studies either did not mention suicide as an exclusion or inclusion criteria (54%) or contained suicide related exclusions that were described without a concrete definition (e.g., suicidal ideation without measure specified or imminent suicidal risk without measure specified) (30%), making strict replication of these studies difficult. Although many of these studies used the same or similar terminology (suicidal ideation or suicidal risk), it is unclear to what extent this may have been interpreted or assessed differently across studies. We think it would be a positive step to call on researchers to more clearly and consistently use concrete criteria to define any such exclusion criteria in future trials.

Additionally, the studies that were found to use suicide measures in defining an exclusion criterion occurred earlier in our data set, suggesting that in recent years researchers have been straying away from clearly defining their terms of exclusion. Outside of the consideration of measures to assess suicide related exclusions, we also found greater use of rates of self-harm and requiring hospitalization exclusions in more recent studies. To our knowledge, there has been little discussion of the value of these specific exclusions. Therefore, it is unclear why these changes have occurred.

We did not find any significant change over time in excluding participants due to depression not being primary. This may indicate that there may still be issues within depression RCT literature that have not changed between 1975 and 2012. Efforts to identify which disorder is primary sometimes consider the order in which the disorders developed, which disorder is more severe symptomatically, or which disorder causes greater functional impairment. In any

case, these are difficult judgments made largely without the aid of a formal assessment suited to the purpose. Because of this, we are not aware of any studies assessing the reliability of evaluators' judgments of which disorder is primary. If whether a diagnosis is primary is important to include in study eligibility determinations, we recommend that researchers develop formal research tools for this purpose.

The only significant relation between year of publication and those falling within the disorder exclusion category were: none, ASPD, and bipolar. The greater use of no disorder related exclusions in earlier years appears to indicate that studies have become more restrictive and started to include less individuals with various psychological disorders other than depression. ASPD's higher frequency as an exclusion in earlier years in our data set contradicts this pattern, with ASPD being included in more recent depression trials, possibly due to a growing consensus surrounding the difficulty of treating ASPD. The increased frequency of bipolar disorder exclusions indicates that more recent studies have more consistently restricted individuals with bipolar disorder from participating in depression research, this is also likely due to a growing recognition of the importance of the distinction between unipolar and bipolar disorders.

Our results also demonstrate that axis I disorders were excluded in 14% of all randomized controlled depression trials. Though this percentage does not appear to be large, previous work done by Stirman, DeRubeis, Crits-Christoph, and Rothman (2005) indicates that individuals with social anxiety disorders who are excluded from RCTs of depression research often do not qualify for other studies looking at their comorbidity. Depression and anxiety are highly comorbid. Therefore, anxiety disorders exclusions have the potential to greatly limit the information provided about whether many patients would benefit from depression treatment. The field could

address this by conducting trials focused on specific comorbidities, testing whether focusing one disorder or the other is a more promising approach. In the absence of such studies, allowing patients with these comorbidities into studies of treatments for depression or anxiety disorders at least allows investigators to go back and examine these comorbidities as moderators of treatment outcome.

Next, we turn to our results from the substance use exclusion category. As previous research has demonstrated, individuals with substance use disorders excluded from depression research were especially likely to be unrepresented elsewhere in the clinical trial literature (Stirman, DeRubeis, Crits-Christoph, & Rothman, 2005). We found that 52% of all studies in our analysis excluded individuals due to some form of substance use condition. The exclusion of such a large number of possible participants demonstrates that much of the previously conducted randomized control trials on adult depression may not generalize to a large portion of the population of interest. There is a need for trials that address such comorbidity.

Of the 173 studies that excluded these individuals, 10% excluded individuals with a previous history of past drug abuse and 9% excluded individuals with a past history of alcohol abuse. This is a sweeping exclusion as these individuals would never be eligible for a study even if their substance use disorders were in remission. An additional 20% of alcohol abuse exclusions did not specify a timeframe for what qualified a participant for exclusion, with the same being true for 23% of all drug abuse exclusions. Open ended exclusion criteria such as this leaves much up to the individual readers' interpretation of the study and makes replicating these studies open to error. Research to address what the best treatment options are for such patients is needed.

## **Limitations**

Our main limitation comes from the range of studies under review. As we chose to use a pre-existing data set to aide us in the task of collecting studies we were limited to analyzing studies from 1975-2012. Since the end of our data collection process the data set in question has been updated through 2018, nearly doubling the number of studies within it for a total of 662 (Cuijpers et. al. 2020). In order to give a fully comprehensive review of exclusion criteria trends in randomized controlled studies of adult depression, studies from more recent years would need to be analyzed.

Another limitation is related to our inclusion of brief reports. Although we would hope that inclusion and exclusion criteria are described no matter how brief a report is, this may not be the case. Due to the brevity of brief reports, authors may have met the length restrictions by shortening their coverage of inclusion/exclusion criteria. If brief reports under report exclusion criteria they used, this would have positively skewed the number of studies that appeared to have no exclusion criteria. It is worth noting that the increased usage of [clinicaltrials.gov](https://clinicaltrials.gov) may help to address this problem moving forward. This website allows investigators to provide detailed inclusion / exclusion criteria and is already seeing widespread adoption.

Additionally, there are important questions that remain to be addressed with our own data set. In future projects, we would like to look into other frequent comorbidity exclusions such as eating disorders, which were coded for within our data set. We also plan to examine journal impact factors of studies and their relations to use of exclusion criteria.

### **Implications**

The results of this study point to some potentially important implications for randomized controlled trials of depression psychotherapy trials. To our knowledge, our project is the first to take a detailed look at the use of various inclusion and exclusion criteria to date. Future research

should continue our work and evaluate more recently published studies on randomized controlled studies of adult depression. We encourage researchers to think carefully about how to operationalize their exclusion criteria so that our confidence in the use of these criteria across trials can be increased.

### **Conclusion**

Between 1975 and 2012 there have been important shifts in the exclusion criteria used in randomized clinical trials of psychotherapies for adults with depression. While many of these shifts are positive, and demonstrate the fields increased knowledge of MDD, others may be seen as problematic for their lack of clear operationalization or their leading to important subgroups of patients who are not adequately represented in the treatment outcome literature. We encourage researchers to think carefully about the inclusion / exclusion criteria they use, so as to not unnecessarily exclude patients who might benefit from a treatment and to use measures of these criteria that can be used consistently across trials and in clinical practice.

### References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, D.C.
- Brière, Frédéric N., Rohde, Paul., Seeley, John R., Klein, Daniel., Lewinsohn, Peter M. (2014). Comorbidity between major depression and alcohol use disorder From adolescence to adulthood. *Comprehensive Psychiatry*, 55(3): 526-533. doi:10.1016/j.comppsy.2013.10.007.
- Cuijpers, P., Karyotaki, E., Ciharova, M., Quero, S., Pineda, B., Munoz, R., ... Figueroa, C. (2020, January 27). A meta-analytic database of randomized trials on psychotherapies for depression. doi: 10.17605/OSF.IO/825C6
- Grant, Bridget F. (1995). Comorbidity between DSM-IV drug use disorders and major depression: Results of a national survey of adults. *Journal of Substance Abuse*, 7, 481-497. doi: [10.1016/0899-3289\(95\)90017-9](https://doi.org/10.1016/0899-3289(95)90017-9)
- Lorenzo-Luaces, L., & Zimmerman, M., & Cuijpers, P. (2018). Are studies of psychotherapies for depression more or less generalizable than studies of antidepressants. *Journal of Affective Disorders*, 234, 8-13. doi: 10.1016/j.jad.2018.02.066
- McHugh, Mary L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3), 276-282.
- Westin, Drew., Novotny, Catherine, M., Thompson-Brenner, Heather. (2004). Empirical Status of Empirically Supported Psychotherapies: Assumptions, Findings, and Reporting in Controlled Clinical Trials. *Psychological Bulletin*, 130, 631-663. DOI: 10.1037/0033-2909.130.4.631.

Wiltsey ., & DeRubeis, R. J., & Crits-Christoph, P., & Rothman, A. (2005). Can the randomized control trial literature generalize to nonrandomized patients. *Journal of Consulting and Clinical Psychology, 73*, 127-135. doi:10.1037/0022-006X.73.1.127.

Wiltsey Stirman, Shannon., DeRubeis, Robert, J., & Crits-Christoph, Paul., Brody, Pamela, E. (2003). Are Samples in Randomized Controlled Trails of Psychotherapy Representative of Community Outpatients? A New Methodology and Initial Findings. *Journal of Consulting and Clinical Psychology, 71*, 963-972. doi:10.1037/0022-006X.

**Table 1: Cohen's Kappa's and Variable Frequencies**

Variable	Kappa	Frequency	Percentage
<b>Suicide Related Exclusions</b>			
None	.91	180	54%
Suicidal Ideation (without measure specified)	.90	90	27%
Suicide Risk (without measure specified)	.98	56	17%
Require Hospitalization Plan	.69	10	3%
Self-Harm	.89	13	4%
Past History	1.00	3	1%
Require Previous Ideation	.80	18	5%
Suicide Measures	.67	3	1%
	.75	22	15%
<b>Disorder Related Exclusions</b>			
Antisocial Personality Disorder	.92	39	12%
Any Bipolar Disorder	.81	135	41%
Borderline Personality Disorder	.76	25	8%
Psychosis	.71	163	49%
Schizotypal	.80	17	5%
Any Personality Disorder	.66	11	3%
Axis I	.94	46	14%
None	.67	99	30%
<b>Depression as Not Primary Exclusions</b>			
Relative to a Specific Diagnosis	.80	30	9%
Diagnosis	.96	126	38%
Multiple Not Primary Exclusions	.91	40	12%
Primary/Principal	.91	31	9%
None	.94	124	37%
<b>Substance Use Exclusions</b>			
None	.90	159	48%
Current Alcohol Abuse	.86	105	32%
Current Drug Abuse	.84	109	33%
Past Alcohol Abuse	.52	15	5%
Past Drug Abuse	.53	18	5%
Alcohol Abuse Not Specified	.70	34	10%
Drug Abuse Not Specified	.68	39	12%

\* Percentage does not equal 100%

Table 2. Differences in Year of Publication for Studies that Did not vs. Did Use Exclusion Criteria

Variable	Not Excluded		Excluded		Z	p
	Median Year	Mean Rank	Median Year	Mean Rank		
<b>Suicide Related Exclusions</b>						
None	2001	171.36	2000	164.19	0.68	.25
Suicidal Ideation (without measure specified)	2001	173.15	2002	165.45	0.64	.26
Suicide Risk (without measure specified)	2001	171.09	2000	148.91	-1.55	.06
Plan	2001	165.97	2005	205.38	1.44	.07
Self-Harm*	2001	166.58	2010	269.33	1.83	.03
Past History Exclusion	2001	167.14	2003	173.89	0.29	.39
Require Previous Ideation	2001	167.75	1998	139.67	-0.50	.31
Suicide Measures *	2001	165.00	2005	206.68	1.87	.03
Require Hospitalization *	2001	169.00	1996	118.65	-1.63	.05
<b>Disorder Related Exclusions</b>						
None*	2002	169.11	1998	142.80	-2.35	.01
Antisocial Personality Disorder*	2001	138.09	1999	164.17	-1.65	.05
Axis I	2000	160.23	2002	182.73	0.79	.21
Bipolar*	2000	152.56	2002	172.63	1.91	.03
Borderline Personality Disorder	2000	161.11	2002	159.68	-0.07	.47
Cluster A	2001	160.79	2006	193.25	0.49	.31
Psychosis	2000	154.11	2001	167.68	-1.31	.09
Schizoaffective	2001	161.35	1998	154.71	-0.29	.39

Table 2 (continued)

Variable	Not Excluded		Excluded		Z	p
	Median Year	Mean Rank	Median Year	Mean Rank		
<b>Depression as Not Primary Exclusions</b>						
None	2000	159.30	2000	162.40	0.29	.39
Diagnosis	2000	163.20	2001	158.74	0.42	.34
Multiple Not Primary Exclusions	2000	160.81	2000	158.35	-0.16	.44
Relative to a Specific Diagnosis	2000	162.19	1999	144.18	-1.02	.15
<b>Substance Use Exclusions</b>						
None	2002	168.86	1999	153.00	-1.53	.06
Current Alcohol Abuse*	2000	173.57	2003	154.90	1.70	.04
Current Drug Abuse*	2000	173.60	2003	154.52	1.75	.04
Past Alcohol Abuse*	2001	162.93	1995	121.57	-1.68	.04
Past Drug Abuse	2001	162.87	1996	129.47	-1.48	.06
Alcohol Abuse Not Specified	2000	158.90	2002	178.69	1.17	.12
Drug Abuse Not Specified	2000	158.63	2002	178.14	1.23	.11