

Codebook

How Does Psychotherapy Improve Youth Depression? Applying Meta-Analytic Structural Equation Modeling to Three Decades of Randomized Trials

ID Variables

Studyid

Type the studyid of the Selected RCT.

TCid

This is the treatment-comparison id to identify the unique treatment-comparison group pair the ES came from, in studyids that have multiple treatment and/or comparison groups that are included.

- If there is one treatment and one control, TCID will always be 1.
- If there is more than one treatment or more than one control group then the TCID will start at 1 for the lowest numbered treatment group or control and then be labeled in ascending order. Examples:
 - Study 7 has 2 treatment and 1 control group
 - 1= Treatment 7.1 & Control 7.3.
 - 2=Treatment 7.2 & Control 7.3.

VarID

Assign a unique VarID to each measure.

For candidate mediators (CMs), use this format: e.g., 720.ma.01, 720.ma.02, 720.mb.01, 720.mc.01, 720.mL2.01

(studyid).(mediator)(category letter “a-e” or “L” for low priority).(2-digit number, 1st one 01, 2nd one 02, 3rd one 03, no particular order required)

For Outcomes, use this format: e.g., 720.o.01, 720.o.02, 720.o.03

(studyid).(outcome).(2-digit number representing the # cms in a category for that study)

Var1id, Var2id

These two ids will tell us which treatment-control comparison, which specific CM, and/or which specific outcome is involved in this ES. For Var1id, list the id of the 1st variable listed in the r(y,x) that is to be recorded for the comparison. For Var2id, list the id of the 2nd variable. If the variable is Treatment Condition, enter “1”. If it is a CM or outcome, then enter the corresponding VarID you coded excluding the studyid.

Var1type, Var1mul, Var2type, Var2mul

For ease of collapsing multiple measures of the same category within each study/treatment-comparison, split each variable id into type (Treatment Condition, Outcome, or CM category) and the multiple number of the measure (e.g., 01, 02). For example:

- If Var1id = 1, then var1type=1, var1mul = NA
- If Var2id = ma.01, then var2type = ma, var2mul = 1
- If Var1id = mL2.03, then var1type=mL2, var1mul = 3
- If Var2id = o.03, then var2type = o, var2mul = 3

Var1time, Var2time

These two items tell us the timepoint of the CM or Outcome. For Var1time, list the time of the 1st variable listed in the r(y,x) that is to be recorded for the comparison. For Var2id, list the id of the 2nd variable.

- Code 0 = if the variable is treatment condition (i.e., no timepoint)
- Code 1 = if the variable is T1 (pretreatment, beginning of treatment)
- Code 2 = if the variable is T2 (midtreatment, during treatment)
- Code 3 = if the variable is T3 (posttreatment, end of treatment)
- Code 4 = if the variable is T4 (follow-up)

Study Characteristics

IPT

- Code 1 = interpersonal psychotherapy, might or might not be a specific version such as *Interpersonal Psychotherapy for Depressed Adolescents (IPT-A)* or *Interpersonal Psychotherapy – Adolescent Skills Training (IPT-AST)*.
- Code 0 = cognitive behavioral therapy, includes interventions where cognitive, or cognitive-behavioral, techniques are a major part of the intervention. Other descriptors include: *problem solving skills training, rational-emotive, restructuring/reframing, self-control training, verbal self-instructional training*.

Author

The author last name of the primary outcome study according to APA citation style.

• Example of APA style for author last name: 1 author: Glass. 2 authors: Smith & Glass. 3 or more authors: Glass et al.

Year

The year of publication of the primary outcome study.

N

Number of participants in the study or the treatment-comparison pair if there are multiple treatment or control conditions.

Mnage

Mean age at time of recruitment. If mean not provided, use the median, and if median is also not provided, use the midpoint of the range of ages of participants included.

- If no ages provided, follow these guidelines by grade: nursery school=3 yo (exclude), preschool=4 yo, K=5 yo, 1st=6 yo, 2nd=7 yo, ... 11th=16 yo, 12th=17 yo, school age=12.
- If mean age provided separately for each treatment condition, take the weighted mean. For example, if there are 10 kids in the control group (mean age = 9.5 years) and 20 kids in the treatment group (mean age = 10.5 years), take $[(10 \times 9.5) + (20 \times 10.5)] / (10 + 20) = 10.2$

Male

Percentage of the sample that is male. Round to 2 decimal places (e.g., 66.67%).

- If not provided, follow these guidelines: Marginal/Few =15%, Some=30%, Majority=60%, Mainly/Predominantly/Primarily=90%.

Ethnicity

Ethcauc	% Caucasian/European American
Eth afr	% African American
Eth lat	% Latino/Hispanic American
Eth asian	% Asian/Pacific Islander American
Eth nat	% Native American
Eth other	% Other/Multi-ethnic
Eth other wh	specify what other/multi-ethnic group
Eth unk	% Unknown

- When computing percentages from frequencies, round to 2 decimal places (e.g., 66.67%).
- If exact percentages are not provided, follow these guidelines: Marginal/Few =15%, Some=30%, Majority=60%, Mainly/Predominantly/Primarily=90%.
- If authors provide some percentage (e.g., 76% white), report what you know about the remaining % (e.g., 24% not white)
- Only use unknown in these cases if the authors report that the ethnicity is unknown for a subsample (e.g., 78% white, 16% African American, 4% Latino/a and 2% unknown)
- Write down a number for each ethnicity category, even when it is 0%.

Dx give

Code whether the participant sample met formal diagnostic criteria for a depressive disorder, including major depressive disorder, persistent depressive disorder (i.e., dysthymia), minor depressive disorder, and depressive disorder-not otherwise specified.

- Code 110 = Yes, all of sample met diagnostic criteria
 - e.g., participants were assessed for diagnosis and included in the study only if they met criteria.
- Code 120 = Yes, some portion of sample met diagnostic criteria
 - e.g., participants were assessed for diagnosis, and some met criteria, but it was not required for study inclusion.
- Code 200 = **None** of sample met diagnostic criteria
 - e.g., participants were assessed for diagnosis and those with a diagnosis were excluded from the study, this is often the case when the aim is to treat subclinical cases or to prevent development of disorder.
- Code 400 = Diagnosis not assessed, mentioned, or reported.

Measure Characteristics

VarName

List the variable name. Type out the full name and put the abbreviation in parentheses. If the author uses a specific name for the measure that is different from the measure's proper name, list the author's name for the measure first and then the measure's proper name. There are often youth- and parent-report, and sometimes teacher-report versions of the same measure. They may

also be adult versions of measures used for adolescents; and there may be “Revised” or “2nd edition” or “3rd edition” versions; short-form vs. long-form versions . Some measures have total scales and separate subscales. Include the most complete/detailed version of the measure name found in the article (usually the first time it is mentioned, often located in the measures/method section of an article), including the specific version for informant and edition, and the specific score or subscale, in full. If there seems to be more than one complete name, include both names, it is better to write down more than less information in this case. Each version/form of the measure/scale should be listed separately in different rows, for example, parent- vs. youth-report versions; short-form vs. long-form versions; total scale score vs. each subscales score computed and/or reported from the same measure; multiple measures are combined or manipulated in any way to form a composite, factor, ratio, count of extreme responses, or other type of score-- include in notes which measures were combined and how they were combined.

Example of listing of total and subscale scores for different informants:

- Revised Children’s Anxiety and Depression Scale (RCADS) total anxiety & depression score
- Revised Children’s Anxiety and Depression Scale (RCADS) total anxiety score
- Revised Children’s Anxiety and Depression Scale (RCADS) total depression score
- Revised Children’s Anxiety and Depression Scale – Parent report (RCADS-P) total anxiety & depression score
- Revised Children’s Anxiety and Depression Scale – Parent report (RCADS-P) total anxiety score
- Revised Children’s Anxiety and Depression Scale – Parent report (RCADS-P) total depression score

Additional Notes

- If the authors reported separate scores for each subscale but did not report a total score, then only list the subscales and not the total. Each subscale should be coded independently from the others (and each one can have a different MedCat).
- If a measure completed by the same informant has subscale scores and a total sum score, the subscale scores and total score do not provide unique information because the same responses go into computing the scores. This applies regardless of whether the subscale scores are all summed or if some are subtracted because they were reversed-scored (i.e., go in the opposite direction from the other subscale scores). Select the broadest variable(s) that fits within each MedCat to code.
 - e.g., If a primary control measure has academic, social, and behavioral subscale, and a total score, select the total score because it is the broadest variable the fits within the primary control coping MedCat for depression studies.
 - e.g., If coping measure has primary control and secondary control subscales and a total adaptive coping score, select the primary control and secondary control subscale scores because each one fits within a MedCat for depression studies.
- If a measure completed by the same informant has subscale scores combined or manipulated in ways other than simple summing/subtracting or averaging (e.g., ratio), the subscale scores and the manipulated score would represent unique information even though the same items went into the subscale scores and manipulated scores.
 - e.g., an explanatory style measure has “negative explanatory scores,” “positive explanatory scores” and a ratio of “negative explanatory scores” to “positive

explanatory scores” then consider, negative scores, positive scores, and the ratio scores as unique information, and thus separate variables to be listed.

VarType

Document the variable type using the following screening codes.

CMs are possible change process or mechanisms through which psychotherapy works to reduce symptoms. CMs usually refer to processes/mechanisms within the youth, or between the youth and someone else, unless otherwise stated. For example, “negative cognitions” and “social skills” refer to negative cognitions and social skills in the youth. If it’s a CM involving others (e.g., parent, therapist), it would be explicitly stated. The authors may or may not use any mediation-related language to refer to them, so you have to use these instructions and your judgment to determine whether a variable is a CM.

- 1 Candidate mediators--high priority (CMH) are CMs that we consider to be high priority because previous research or theory have indicated that these are likely to explain how and why certain EBPs work. CMH will differ by Big 4 target problem, though there may be some overlap in CMH across problem area. We will definitely code and include these in our analyses. If a CM seems equally like it could be a CMH or a CML, prioritize CMH coding. If the CM is a combination of CMH MedCats, code as CMH.
- 2 Candidate mediators--low priority (CML) are CMs that we consider to be low priority because these have not been studied much with EBPs among youths, and/or are not considered in theory to play a substantial role in how EBPs work. We will code these but may or may not include these in our analyses, depending on how much data on these we can get. Or we could consider them for inclusion in future analyses. There’s also a chance that what we considered CML ends up having lots of data and we may “upgrade” them to CMH status.
- 3 Outcomes are measures of youth depression symptoms.
- 4 Outcomes- are measures of youth symptoms that are clearly narrower than depression symptoms, for example, a measure of suicidal ideation.
- 5 Outcomes+ are measures of youth symptoms that are clearly broader than depression symptoms but narrower than overall symptoms, for example, a measure of internalizing symptoms.

Other measures that do not fall into the above categories include outcomes of some other target problems, overall/general symptoms, non-outcomes of the youth such as general youth functioning and environmental impact of the treatment on parents or siblings that are not related to the youth or to their relationship with the youth (e.g., parental marital conflict, stress experienced by parent not related to parenting the child). This also includes variables that are not identical across participants (i.e., free-response, open-ended, idiographic type answers).

- Code CMH=1, CML=2, outcome=3, outcome-=4, outcome+=5.

See list below for specific examples of CMs and outcomes. If you don’t find the measure you are coding, read over several examples to get an idea of what should come under Outcomes and each CM to make your best judgment.

1 Candidate mediators--high priority (CMH)

a: Negative Cognitions

- Extent of negative thinking; if it's a positive thinking measure, make a judgment about whether the authors conceptualize the positive thinking as the opposite of negative

thinking (i.e., youths who score high on one end of the measure are considered positive thinkers and youths who score on the low end of the measure are considered negative thinkers or vice versa). If the authors do so, then code positive thinking here. If authors conceptualize positive thinking as a way to displace negative thoughts, code as Cognitive Reframing.

- Cognitive distortions, jumping to negative conclusions
- Irrational, unrealistic beliefs/expectations
- Depressogenic thoughts/schemas
- Dysfunctional attitudes
- Pessimistic explanatory/attributional style
- Negative self-concept/self-esteem/self-worth, negative beliefs/views about oneself
- External locus of control, low perceived sense of control
- Depressive/negative self-talk
- Negative memories
- Negative interpretations
- Negative information-processing bias
- Children's Negative Cognitive Error Questionnaire (CNCEQ)
- Automatic Thoughts Questionnaire (ATQ)
- Dysfunctional Attitude Scale (DAS)
- Beck Hopelessness Scale (BHS)
- Rumination/worry--repetitively think about the causes, situational factors, and consequences of one's negative emotional experience
- Perceived uncontrollability of depression
- Self-blame for depression
- Meta-cognitive awareness
- Piers-Harris Children's Self-Concept Scale (PHCSCS)
- Perfectionism

b: Pleasant Activities

- Engagement in activities one enjoys
- Activity scheduling
- Pleasant Events Schedule (PES)
- Include physical activities, hobbies, relaxation, but not social activities

c: Problem-Solving

- Skill or effort in taking steps to change external, objective conditions, could be interpersonal conflicts or other kinds of stressors or problems
- Primary control coping, active coping
- Conflict resolution skills
- Generating solutions
- List of Social Situations Problems
- Ask others for help to solve a problem

d: Reframing

- Skill or effort in changing one's internal, subjective thinking, to deal with stressors, negative thoughts, problems, thus adapting to external, objective conditions
- Secondary control coping

- Distraction
- Thinking positive thoughts or something else in place of negative thoughts, or with the intention to improve mood; but if authors conceptualize positive thoughts as a measure or more vs. less negative thoughts, code as Negative Cognitions.
- Finding the silver lining
- Coming up with alternative explanations
- Acceptance
- Distancing from negative thoughts
- Identifying/being aware of negative thoughts

e: Social Engagement

- Social skills
- Participation in social activities
- Positive self-presentation, present self as friendly and sociable
- Social adjustment
- Quality of interpersonal relationships
- Social support
- Loneliness
- Sociometric ratings of likability/popularity
- Prosocial behavior
- Positive behavior towards peers
- Matson Evaluation of Social Skills with Youngsters (MESSY)
- Note: If 2/3 of items are general social skills/relationships or with peers, and 1/3 with family, measure would still fit here.

f: Combination of 2 or more CMH MedCats

2 Candidate mediators--low priority (CML)

L2 Family problematic relationships

- Any family relationship problem, including dysfunction, lack of cohesion
- Parent/family expressed emotion (criticism, hostility, or emotional overinvolvement of family members)
- Communication problems
- Conflicts, fights
- Relationship quality

L29 Avoidance

- avoidant coping
- wishful thinking
- denial/avoidance of the problem

Note: Only L2 and L29 had sufficient data to be examined in the present meta-analysis, other CML codes listed below were not examined in the present meta-analysis.

- L1 Readiness to change
- L3 Child-therapist therapeutic alliance and therapeutic relationship. Alliance--the collaborative and affective relationship between the therapist and the client, includes extent of agreement between therapist and client about goals of therapy, techniques that

will be implemented to achieve the goals, (i.e., tasks), and the bond between therapist and client (e.g., mutual trust, liking, respect, caring). Other therapeutic relationship variables: Experiential congruence: that the degree to which client and therapist agree on how they experience their session from a relational perspective. Expressive attunement: the quality of communication between the client and the therapist, including therapist expressiveness, patient expressiveness, the patient's empathic understanding in therapy, and communicative rapport. Affective attitude: feelings that the client and therapist have towards one another. Includes therapist affirmative behavior, patient affirmative behavior towards the therapist, the joint affirmative behavior of both therapist and client.

- L4 Parent-therapist therapeutic alliance/relationship.
- L5 Understanding and finding meaning in loss and change
- L6 Understanding affect and their link to relationships
- L7 Insight, self-understanding, self-understanding of interpersonal patterns
- L8 Therapist cognitive strategies, therapist adherence to and competence in the treatment protocol of CBT or CT, sometimes called treatment fidelity/integrity for CBT, therapist knowledge of CBT
- L9 Therapist behavioral strategies, therapist adherence to and competence in the treatment protocol of BT, sometimes called treatment fidelity/integrity for BT, therapist knowledge of BT
- L10 Therapist psychodynamic strategies, therapist adherence to and competence in psychodynamic therapy protocol, sometimes called treatment fidelity/integrity for psychodynamic therapy, therapist knowledge of psychodynamic therapy; specific strategies such as Dynamic interpretations, transference interpretations
- L11 Therapist family systems strategies, therapist adherence to and competence in family or family systems therapy protocols, sometimes called treatment fidelity/integrity family systems therapy, therapist knowledge of family therapy
- L12 Therapist client-centered strategies, therapist adherence to and competence in client-centered therapy protocol, sometimes called treatment fidelity/integrity for client centered therapy, therapist knowledge of client-centered therapy
- L13 Therapist interpersonal therapy strategies, therapist adherence to and competence in the treatment protocol of IPT sometimes called treatment fidelity/integrity for IPT, therapist knowledge of IPT
- L14 Therapist flexibility
- L15 Level of collaboration between therapist and client, directiveness/controlling behavior of therapist (opposite of collaboration)
- L16 Therapist role engagement and credibility--specific to the the therapist
- L17 Youth role engagement and motivation, participant, involvement in therapy, treatment/homework completion/compliance, knowledge of therapy content. We will NOT include attendance as a CML.
- L18 Parent role engagement and motivation, participant, involvement in therapy, treatment/homework completion/compliance, knowledge of therapy content. We will NOT include attendance.
- L19 Treatment credibility (specific to treatment type), preference, expectancy (extent to which one expects to benefit from treatment); include attitudes towards treatment, including towards therapy, medication etc.

- L20 Mindfulness
- L21 Self-compassion
- L22 Psychological flexibility: the ability to fully embrace thoughts, feelings and experiences in the present moment without avoidance and persisting or altering behaviour to be consistent with goals and values
- L23 Cognitive and emotional reactivity: the extent to which a mild state of distress coupled with stress reactivates negative thinking and emotional patterns, putting individuals at risk of a depressive episode
- L24 Group cohesiveness of treatment group
- L25 Autobiographic memory specificity: the ability to retrieve memories of specific personal events that happened at particular times and locations.
- L26 Meta-awareness and decentering
- L27 Expression of affect/emotion--this is a more general term for expressing feelings, talking about it, communicating about it in writing or other creative outlets
- L28 Youth, family, or therapist satisfaction with treatment protocol in the study or with the services received that were a part of the treatment condition or a direction result of the treatment condition (e.g., participants received usual care, or directed to access specialized services, as part of the study condition--satisfaction can be for these services as well)
- L99 Other candidate mediator

3 Outcomes

- Major depressive disorder
- Dysthymic disorder/persistent depressive disorder
- Minor depression
- Adjustment disorder with depressed mood
- Beck Depression Inventory (BDI)
- affective problems DSM-oriented scale
- Child Behavior Checklist (CBCL) withdrawn/depressed subscale
- Child Behavior Checklist (CBCL) affective problems DSM-oriented scale
- Youth Self-Report (YSR) withdrawn/depressed subscale
- Youth Self-Report (YSR) affective problems DSM-oriented scale
- Teacher Report Form (TRF) withdrawn/depressed subscale
- Teacher Report Form (TRF) affective problems DSM-oriented scale
- Child Depression Inventory (CDI)
- Children's Depression Rating Scale (CDRS)
- Center for Epidemiological Studies Depression Scale for Children (CES-DC)
- Hamilton Rating Scale for Depression (HRSD, Ham-D)
- Revised Children's Anxiety and Depression Scale (RCADS) depression subscale
- Mood and Feelings Questionnaire
- Patient Health Questionnaire 9

4 Outcomes-

- Subscales of measures in Outcome
- Depressed mood, sadness, unhappiness

- Irritability
- Anhedonia, loss of interest or pleasure
- Weight loss or gain, decrease or increase in appetite
- Insomnia or hypersomnia
- Psychomotor retardation or agitation
- Fatigue or loss of energy
- Worthlessness
- Guilt
- Brooding about the past
- Diminished ability to think or concentrate
- Indecisiveness
- Thoughts of death, suicidal ideation
- Suicide plan or attempt
- Feelings of inadequacy
- Feelings of pessimism
- Social withdrawal in the context of depression (not anxiety)
- Decreased activity
- Decreased effectiveness or productivity

5 Outcomes+

- Internalizing/overcontrolled problems
- Depression and anxiety
- Neurosis
- Pain
- Hypochondriasis
- Physical complaints (e.g., stomachaches)
- Somatization
- Child Behavior Checklist (CBCL) internalizing scale
- Child Behavior Checklist (CBCL) anxious/depressed subscales
- Youth Self-Report (YSR) internalizing scale
- Youth Self-Report (YSR) anxious/depressed subscales
- Youth Self-Report (YSR) somatic subscales
- Teacher Report Form (TRF) internalizing scale
- Teacher Report Form (TRF) anxious/depressed subscales
- Child Behavior Checklist (CBCL) somatic subscales
- Revised Children's Anxiety and Depression Scale (RCADS) total anxiety & depression score
- Strengths and Difficulties Questionnaire (SDQ) emotional symptoms subscale
- Positive and Negative Affect Schedule (PANAS) - Negative affect

MedCat

Code the mediator category for each CM. Most of the time, we would see the title of the measure, the contents of the measure (actual questions, description in Measure section of article, validation study description of measure), and the RCT authors' discussion of what the measure is supposed to do (e.g., in intro, results, discussion sections) all match up, and it is clear which category the CM comes under. However, there are some few tricky situations. For example, the 3 pieces match up but the measure seems to encompass two or three categories, or the 3 pieces do not match up and could come under one or another category depending on which piece we prioritize. We would prioritize the contents of the measure over the title of the measure because sometimes titles don't reflect what the measure is about, and over the RCT authors' discussion of the measure because the RCT authors sometimes use terminology/wording quite loosely (e.g., "coping" is used to describe many different measures with different content). So if the measure contents clearly come under one MedCat, code under that MedCat even if the title or RCT author may use wording that seems to put the measure in another MedCat. If the measure contents don't come clearly under one MedCat, but the RCT authors' discussion/interpretation of the measure suggests it come under one, then go with how the authors discuss the measure. Occasionally, you might see a same-RCT article do a factor analysis of items on a single measure, or total score or subscale scores of multiple measures--you can use this as clues as to how the RCT authors interpret the variable. For example, there was a "problem-solving orientation" measure that could be considered either negative cognitions or problem solving and the factor analysis results state that this measure loaded onto a solution-focused thinking factor, then code as problem solving. The $\frac{2}{3}$ rule can help you make a decision if a measure contains content that falls under two MedCats--if at least $\frac{2}{3}$ of the measure falls within one category we'll count it in that category even if there's a small amount of content from another category. If a measure clearly encompasses two MedCats equally (and has no subscale scores for each MedCat), create a new code under CML and code it there. We will decide later whether to merge MedCats, create a new MedCat, or do something else.

- Code Category a=1, b=2, c=3, d=4, e=5
- For combination of CMH, code 6 and note the CMHs combined. Code here ONLY if CM cannot be coded under the 5 CMH categories.
- For CMLs see list above
 - e.g., coding a measure of youth-therapist alliance for depression would be coded under medcat as L3
 - e.g., coding a measure that doesn't fall into any category would be coded as L99 when first completing Variable tab for a TP; later we may recode those under L99.
- For Outcomes, code 99.

Source

Source of rating reflects who provides the actual number that is analyzed (i.e., rater, reporter, informant, observer), regardless of who is the subject of the rating. Code according to the authors' report of who coded the measure(s) or the actual measure instructions/validation study information on who is meant to complete the measure (e.g., the CBCL is designed to be completed by parents, the YSR is meant to be completed by youths about themselves). Other studies that are NOT the Selected RCT that happened to use the same measure could be used for coding if it provides information on who is meant to complete the measure, but if the measure is

meant to be completed by any of multiple informants (e.g., family stress measure that is designed to be completed by any member of the family) and the other studies happened to use parents to complete the measure, we cannot assume that the Selected RCT also used parents to complete the measure.

- 1 Target youth: This includes measures where the target youth reports on their own problems or on someone else's problems through self-report measures or interviews (e.g., *affect, behavior, competence, self-esteem, number times they wet bed, personality, interviews that consist of reading a self-report questionnaire where the interviewer just records the subjects' responses*). The youth's responses could be summed up or averaged by a researcher, or have various manipulations performed on them by a computer program (e.g., factor analysis, principal components analysis, composite)--as long as there's no room for judgment by anyone else, the source would still be the child. If there is room for a scorer, judge, rater, coder, observer etc. to make a judgment about the responses to come up with a score or number, then that person would be the source instead of the youth (see 7 Researcher for more details). This applies to the other sources as well.
- 2 Caregiver: Main caregiver, usually parents, legal guardians, foster parents, provide the rating. This includes parent providing other-report measures (e.g. *recording the number of days that the child wet the bed*), and parent providing self-report of their own problems/issues. If two separate versions are measured, e.g. mother-rated survey AND father-rated survey, consider them as you would any other measures, i.e. coded in 2 different rows.
- 3 Sibling: Siblings of the target youth provide the rating about themselves or about someone else.
- 4 Peer: Ratings made by peers and sociometric measures.
- 5 Teacher: Ratings made by teachers or individuals who play similar roles to teachers (e.g., camp counselors, milieu workers, "big buddies", probation officers, etc.).
- 6 Therapist: Therapists (i.e., the individuals administering the intervention) provide the rating about themselves or about someone else. When the "therapist" is a teacher, parent, or peer, they are coded in that source category rather than here.
- 7 Researcher: Ratings made by a member of the research team, include research assistants, study clinicians who are not the youth's therapist, independent evaluators, through observation, counting behaviors, and administering tests. If the ratings are made based on interview, the rater must be making a judgment based on the response, examples include GAF, CGI, C-GAS. Most performance-based (instead of subjective-report) measures will also be coded here (e.g., *height, weight or percent body fat, holding capacity of bladder, heart rate/pulse, galvanic skin response, performance on intelligence / achievement test, MFFT, approach / avoidance / other behaviors in an "exposure" test for anxiety*). In these cases, the researcher is making a judgment about what to code/count as a target behavior, what responses is correct or incorrect on a test, what reading is showing on a heart rate monitor (even though the room for judgment may be small on a machine reading, the researcher is still the one making the judgment). Usually, there would be some details provided on the reliability of the judgment by different coders/raters, and the training they received for interview/observation measures.
- 8 Life-event data: When the data come from a source independent of the research project. Common examples that generally fall here include: *arrests, attendance records, clinic service records of services used outside of the treatment study (if it comes from youth or caregiver report, code 1 or 2), grades, measures given to an entire class (e.g. achievement tests), milieu fines, school dropout, school placement, suspensions, urine analysis / drug screen (if it comes*

from a source independent of the research project like probation department; if it was done for the research project, it falls under 7 Researcher). Note that a Life Events Checklist completed by parent or child is still just other or self report – it does not count as life event data from an independent source.

- 9 Unknown: Authors did not report the source, the actual measure/validation studies/other studies that are not Selected RCTs report that the measure was designed to be completed by any of several informants, and authors did not reply about the source when contacted. For this code do not make an assumption that the measure was coded by a particular source just because the other measures listed next to them were coded by that source.
- When there are multiple sources for a measure or for a factor score combining multiple measures, code the smallest number.
- Code youth=1, caregiver=2, sibling=3, peer=4, teacher=5, therapist=6, researcher=7, life-event=8, unknown=9.

OppDir

To combine data across measures and across studies, we have to assign an “expected direction” and correct any measures that are in the “opposite direction.”

Outcomes: For most outcome measures, a higher score means more severe symptoms--this is the expected direction. However, some outcome measures may have higher score to mean better functioning, or greater improvement--this is in the opposite direction and we will need to reverse scores from this measure before combining with other measures that are in the expected direction.

CMs: For CMs, the expected direction is going to vary from CM to CM. We will consider the expected direction to be reflected by the CM category name. For example, Negative Cognitions high score meaning more negative cognitions will be the expected direction. Any Negative Cognition measure in which higher score means less negative cognitions or more positive cognitions will be considered as being in the opposite direction. Another example, Social Engagement high score would mean more/better social engagement is the expected direction. Any social engagement-related measure where high score means worse social engagement will be considered as being in the opposite direction. DO not code OppDir for L99 because these CMs don't have their own category; if we classify L99s and give them formal MedCats later, then we will code the OppDir.

This is critical for computing accurate effect sizes, so follow the general strategies below for coding variable characteristics. For example, you are coding the CM called “Children’s Negative Cognitive Errors Questionnaire,” which you coded as MedCat Negative Cognitions.

1. From the VarName, it seems like higher score may be more negative cognitive errors. This is in the same direction as of MedCat Negative Cognitions. Looks like this measure is in the expected direction.
2. Check the method/measures section, to confirm your guess. It turns out the authors reported that lower scores reflect greater cognitive distortion. Looks like this measure is actually in the opposite direction.
3. Next, read the results, tables, figures and discussion to see whether the numbers and description of findings are consistent with what you think. When all this information that you’ve checked is consistent, you can confidently code “opposite direction” for this measure.

4. If the articles don't contain enough information, check relevant measures and validation studies.
5. If there is simply no clear evidence for OppDir, err on the side of coloring orange and emailing the RCT authors to ask to avoid coding an effect size in the reverse direction from the true effect size. If the RCT authors don't respond, email the measure validation study authors.

- Code opposite direction=1, expected direction=0.

T1

Indicate whether this variable was measured at pretreatment (i.e., before treatment started, baseline) or not (regardless of whether they provided the data or not). Occasionally, variables are measured at the beginning of treatment (session 1) rather than at pretreatment, especially for change process variables that can only occur in the context of treatment. If you have any variables NOT measured at pretreatment/session 1 but measured "early" in treatment (session 2 onwards) let me know, we'll need to make a case-by-case decision on how to code that.

- Code T1 yes=1, no=0.

T2

Indicate whether this variable was measured at midtreatment or not (regardless of whether they provided the data or not). We will consider as midtreatment anything measured after session 1 and before the last session, including measurements taken multiple times during treatment (e.g., several therapy sessions sampled for coding for each youth). Note that measures of things that occur during the treatment itself would come here (e.g., treatment adherence, treatment strategies, therapeutic alliance) because they do not come before or after the treatment--even if the coding itself is done after the treatment ends, the tapes or transcripts of treatment sessions coded were collected during the treatment itself. Include in the notes any variables for which there are more than 1 midtreatment measurement (e.g., session 3, 6, 9, 12). We'll pick the measure that is closest to the mean/median/midpoint within each youth, and then averaged across youths. For example, if a measure was given at pretreatment, and then every second session until the last session in a 12-session treatment, then there would be 5 midtreatment measurements at session 2, 4, 6, 8, and 10 (0 and 12 would be pre and post). We'll pick the session 6 measurement in this case.

- Code yes=1, no=0.

T3

Indicate whether this variable was measured at posttreatment (i.e., shortly after treatment ended) or not (regardless of whether they provided the data or not). Sometimes authors do not refer to a "posttreatment" timepoint; instead they have several different assessment points they call "follow-ups" at 3, 6, 9, months, 1 year etc. Select the timepoint that occurs earliest and < 6 months (24 weeks) after the end of treatment on average in the sample. Thus for a treatment that's 12 weeks long, the posttreatment timepoint would be the "3-month follow-up." If posttreatment assessment was conducted on average 2 months after the end of treatment, but the range across the sample was 0.5 to 7 months, we would still count this as a posttreatment timepoint. It's possible that some studies are not interested in improvement after the end of treatment, but rather that the onset of a disorder is prevented 1 year after the end of treatment; thus if the earliest assessment after the end of treatment is 6 months or more on average, we will

count this timepoint not as posttreatment but as follow-up. Occasionally, variables are measured at the end of treatment (last session) rather than at posttreatment, especially for change process variables that can only occur in the context of treatment. If you have any variables NOT measured at pretreatment/session 1 but measured “late” in treatment (last few sessions) let me know, we’ll need to make a case-by-case decision on how to code that.

- Code yes=1, no=0.

T4

Indicate whether this variable was measured at follow-up (i.e., some time after posttreatment assessment) or not (regardless of whether they provided the data or not). If the authors call the posttreatment assessment “follow-up” and there’s no other assessment after that, do not count the study as having a follow-up. In the rare case where the earliest assessment after the end of treatment is 6 months or longer on average after the end of treatment, this assessment will be counted as a follow-up and there is no posttreatment assessment. The variable needs to be measured at follow-up for at least one treatment condition and at least one control condition where there is no switching of conditions before the follow-up assessment to be considered as being measured at follow-up. For example, an RCT compared a treatment group and a waitlist control, at pretreatment, posttreatment, then switched the waitlist over to the treatment, and after treatment did another assessment for both groups. The last assessment is a follow-up for the treatment group but has become a posttreatment assessment for the control that switched over to treatment, thus this last assessment will NOT be considered as a follow-up assessment for both groups, and T4 should be coded as 0.

- Code yes=1, no=0.

Effect Size Data

NumOppDir

Indicate the number of opposite direction variables in this ES. Here, indicate whether 0, 1 or 2 variables in this ES are in the opposite direction. For CMs and Outcomes, refer to your coding of OppDir. Treatment/control condition has no direction so in the case you look at an ES of treatment effect on a CM or outcome, you will only need to check whether the CM or outcome is in opposite direction.

- Code 0 = all measures are in the expected direction, none are OppDir
- Code 1 = 1 measure in the variable pair is OppDir, the other is in the expected direction or is treatment condition. We will later use syntax to reverse these ES.
- Code 2 = 2 measures of the variable pair are OppDir.

Aside from the above rules, only if coding for a correlation that involves a variable that does not exist in the study, which will therefore have a NA for r or for mean_t through n_c variables, always put a 0 for opposite direction since we will not reverse code it.

ESmethod

Indicate the method by which you derived the ES if you did not code means, SDs, or r , and n .

0 NS: If the only usable information we have, in the form of words or statistics that we cannot convert, is that there is no significant effect/association between two variables, code ESmethod=0 and $r=0$.

99 Uncodable: No data were reported for that ES, or the data reported are not convertible because the data involved the influence of variables in addition to the two variables that make up the ES. Describe the issue briefly and enter ESmethod = 99.

Other methods

- 2 Unstandardized regression coefficient (b)
- 3 $k \times j$ frequency table
- 12 Independent t-test statistic
- 13 One-way ANOVA for two groups, F-test statistic
- 14 Independent t-test p-value
- 15 Means and standard errors
- 16 Point-biserial correlation
- 17 Point-biserial correlation p-value:
- 18 Frequency distribution
- 19 Unstandardized regression coefficient (b) and SD of Outcome/CM: Bivariate
- 20 Standardized regression coefficient (β) and SD of Outcome/CM
- 21 Means and full-sample SD
- 22 Mean gain, SD of gain, pre-post r
- 23 Mean gain, pre/mid/post SD, paired t-test
- 24 Mean gain, pre/mid/post SD, pre-post r
- 26 One-way ANOVA for three or more groups, F-test statistic
- 27 ANCOVA for two groups, MS error, covariate-Outcome/CM correlation, means
- 28 Two-way ANOVA for two groups, F-test statistic, df, MS error

r

For cells involving mediator-outcome or mediator-mediator variable pairs, find the simple correlation r (i.e., Pearson's product-moment correlation between two continuous variables) reported in the article and plug it into the ES tab. Because mediators and outcomes that we are including are continuous variables, the most appropriate ES to code is r , but authors may or may not report this. If not found in the article, highlight this cell in orange, find other statistics that represent the association between the two variables of interest and detail them in the notes.

If an article reports this information for the participants in multiple treatment groups combined v. control and does not provide it for any individual treatment and control comparisons then create a tab for combined treatment and code the correlations given for the full sample if we were interested in all control and treatment groups. If the sample for the correlations includes a group condition we were not planning to include as tx or control bring it up to the group.

If coding for a correlation that is the relationship between a variable in the study against itself at the same time point, code $r = 1$ because a variable is perfectly correlated with itself. Even if this variable at this timepoint is missing by design or by report, code 1 because the metaSEM needs the V cells to be 1.

If coding for a correlation that involves a variable that does not exist in the study, then code $r = \text{NA}$.

DO NOT round decimal places but use the exact one from the cell in the correlation table output.

n_r

Enter the sample size used to calculate the r only if you record the r . Note that the sample size depends on the sample size of both variables involved, r can be computed only for youths who have both variables available at that timepoint. First try to find the specific sample size for this variable measured at their specific timepoints, usually reported close to the r . If not available, use your best estimate, which could be the total sample size at the timepoint(s) the variables were measured, whichever is the smaller number (e.g., if one variable is T1 and the other variable is T3, take the smaller sample size which is often at the later timepoint), often shown in the flowchart of participants at various stages of the RCT. Look for author's description in words of any participants not completing timepoint assessments and subtract them from the total sample size if not already subtracted. If not found in the article, highlight this cell in orange and leave blank--fill in later when author provides statistic together with the n . If author does not provide information, make an assumption here about which n to use based on the n of similar measures administered together.

If coding for a correlation that involves a variable that does not exist in the study, the r_n will then be NA.

If coding for a correlation that is the relationship between a variable in the study against itself at the same time point, code for r_n the total number of subjects from both treatment and control that filled out that measure ($n_t + n_c$ for that measure at that timepoint). **If there are 3 of the same category of CMs/Outcomes at that timepoint, then take the mean of ($n_t + n_c$) for all 3 same category-CMs/Outcomes by using the excel AVERAGE function, copy, and then "paste value" only so that all the decimal places of the mean n will go into the cell but the function will not be in the cell (which might help with loading files into the statistical program).**

To find the n for the first row, you should first look at:

1. The total n ($n_t + n_c$) given for the study.

If this is not available, you should next look at:

2. The total n ($n_t + n_c$) at randomization. If 80% of it corresponds with the n from the largest measure ($n_t + n_c$ for this measure), then you can use it.

If this is not available, next look at:

3. The largest n ($n_t + n_c$) for a measure.

If none of these are available, or if anything comes up that you are unsure about, feel free to bring it up for discussion.

mean_t, mean_c

For cells involving treatment/control-outcome or treatment/control-CM variable pairs, find the mean score of the CM or Outcome for the treatment group and for the control group respectively and plug them into the ES tab. Because treatment/control condition is a dichotomous variable (1=treatment, 0=control), the data will likely be reported as means and SDs for each group, which we can convert into r . If not found in the article, highlight the cell(s) in orange, find other statistics that represent the association between the two variables of interest and detail them in the notes.

DO NOT round decimal places but use the exact one from the cell in the correlation table output.

SD_t, SD_c

Enter the standard deviation (SD) in CM or Outcome scores for the treatment group and for the control group respectively. If standard errors (SE) are reported, use this formula $SD = SE \times \sqrt{N}$ --be sure to use the correct N from the same group at that timepoint when the SE was computed. If not found in the article, highlight the cell(s) in orange, find other statistics that represent the association between the two variables of interest and detail them in the notes.

DO NOT round decimal places but use the exact one from the cell in the correlation table output.

n_t, n_c

Enter the sample size of the treatment group and of the control group respectively that was used in the computation of the means and SDs for this particular variable. If the sample sizes are not reported for each variable, but are reported for the study as a whole at each timepoint, you can use the whole-study sample sizes for the specific timepoint that you are coding ES for. Look for author's description in words of any participants not completing timepoint assessments and subtract them from the total sample size if not already subtracted. If you are unsure about sample sizes because the authors used some kind of imputation technique, or report both imputed and raw data, document in the notes and bring up for discussion. We usually would go for the raw data and/or lower number because this is more conservative. If not found in the article, highlight the cell(s) in orange. leave blank--fill in later when author provides statistic together with the n. If author does not provide information, make an assumption here about which n to use based on the n of similar measures administered together.