Table 1. Domains for determining risk of bias in systematic reviews of psychotherapy outcome studies [partially adapted from Higgins et al. (2011)].

Domain	**Sample criteria for low risk of bias**	**Sample criteria for high risk of bias**
Sequence generation: *Was the allocation sequence adequately generated?*	The investigators describe a random component in the sequence generation process such as referring to a random number table, using a computer random number generator, or coin toss	The investigators describe a non-random component in the sequence generation process
Allocation concealment: *Was allocation adequately concealed?*	Participants and investigators enrolling participants could not foresee assignment because adequate methods (e.g., central allocation, sequentially numbered envelopes)	Participants or investigators enrolling participants could possibly foresee assignments and thus introduce selection bias, such as allocation based on an open random allocation schedule (e.g. a list of random numbers) or allocation based on unconcealed or non-random factors
Blinding of study personnel and outcome assessors: *Was knowledge of the allocated interventions adequately prevented during the study?*	Any of the following: • No blinding, but the review authors judge that the outcome and the outcome measurement are not likely to be influenced by lack of blinding • Blinding of key study	Any of the following: • No blinding or incomplete blinding, and the outcome or outcome measurement is likely to be influenced by lack of blinding • Blinding of key personnel was attempted, but likely

	personnel ensured, and unlikely that the blinding could have been broken • Some key study personnel were not blinded, but outcome assessment was blinded and the non-blinding of these personnel was unlikely to introduce bias	that the blinding could have been broken • Some key study personnel were not blinded, and the non-blinding of these personnel was likely to introduce bias
Blinding of participants: *Was knowledge of the allocated interventions adequately prevented during the study?*	Any one of the following: • No blinding, but the review authors judge that the outcome and the outcome measurement are not likely to be influenced by lack of blinding • Blinding of participants ensured, and unlikely that the blinding could have been broken • Participants were not blinded, but outcome assessment was blinded and the non-blinding of participants unlikely to introduce bias	Any one of the following: • No blinding or incomplete blinding, and the outcome or outcome measurement is likely to be influenced by lack of blinding • Blinding of study participants was attempted, but likely that the blinding could have been broken • Participants were not blinded, and the non-blinding of participants was likely to introduce bias
Incomplete outcome data: *Were incomplete outcome data adequately addressed?*	Any one of the following: • No missing outcome data • Reasons for missing outcome data unlikely to	Any one of the following: • Reason for missing outcome data likely to be related to true outcome,

	be related to true outcome • Missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups • For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a clinically relevant impact on the intervention effect estimate • For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a clinically relevant impact on observed effect size • Missing data have been imputed using appropriate methods	with either imbalance in numbers or reasons for missing data across intervention groups • For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce clinically relevant bias in intervention effect estimate • For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce clinically relevant bias in observed effect size • 'As-treated' analysis done with substantial departure of the intervention received from that assigned at randomization • Potentially inappropriate application of simple imputation
Selective outcome reporting: *Are reports of the study free of suggestion of selective*	Any of the following: • The study protocol is available and all of the	Any one of the following: • Not all of the study's pre-specified primary

outcome reporting?	study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way • The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified	outcomes have been reported • One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified; • One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect) • One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis • The study report fails to include results for a key outcome that would be expected to have been reported for such a study
Treatment fidelity: *Was the treatment implemented as intended?*	All of the following: • Therapists had adequate qualifications and training to provide the study treatment • A publicly-available	Any of the following: • Therapists were not adequately qualified or trained to provide the study treatment • No publicly-available

	treatment manual was used • Adherence to the treatment protocol was monitored and judged to be adequate	treatment manual was used • Adherence to the treatment protocol was either not monitored, or was monitored and is judged to have been inadequate
Other potential threats to validity: *Was the study apparently free of other problems that could put it at a risk of bias?*	The study appears to be free of other sources of bias	There is at least one important risk of bias. For example, the study: • Had a potential source of bias related to the specific study design used • Stopped early due to some data-dependent process (including a formal-stopping rule) • Had extreme baseline imbalance • Has been claimed to have been fraudulent • Had some other problem, including clear financial conflict of interest

Table 2. Summary assessments of risk of bias [adapted from Higgins and Green (2008)]

Risk of bias	**Interpretation**	**Within a study**	**Across studies**
Low risk of bias	Plausible bias unlikely to seriously alter the results	Low risk of bias for all key domains	Most information is from studies at low risk of bias
Unclear risk of bias	Plausible bias that raises some doubt about the results	Unclear risk of bias for one or more key domains	Most information is from studies at low or unclear risk of bias
High risk of bias	Plausible bias that seriously weakens confidence in the results	High risk of bias for one or more key domains	The proportion of information from studies at high risk of bias is sufficient to affect the interpretation of the results

Table 3. Scoring criteria for the Assessment of Multiple Systematic Reviews (AMSTAR) system (Shea et al., 2007)

1. Was an 'a priori' design provided? The research question and inclusion criteria should be established before the conduct of the review.	Yes No Can't answer Not applicable
2. Was there duplicate study selection and data extraction? There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.	Yes No Can't answer Not applicable
3. Was a comprehensive literature search performed? At least two electronic sources should be searched. The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.	Yes No Can't answer Not applicable
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion? The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.	Yes No Can't answer Not applicable
5. Was a list of studies (included and excluded) provided? A list of included and excluded studies should be provided.	Yes No Can't answer Not applicable
6. Were the characteristics of the included studies provided? In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.	Yes No Can't answer Not applicable

7. Was the scientific quality of the included studies assessed and documented? 'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.	Yes No Can't answer Not applicable
8. Was the scientific quality of the included studies used appropriately in formulating conclusions? The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.	Yes No Can't answer Not applicable
9. Were the methods used to combine the findings of studies appropriate? For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?).	Yes No Can't answer Not applicable
10. Was the likelihood of publication bias assessed? An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).	Yes No Can't answer Not applicable
11. Was the conflict of interest stated? Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.	Yes No Can't answer Not applicable

References

Higgins, J. P., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., . . . Cochrane Statistical Methods, G. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal, 343*, d5928. doi: 10.1136/bmj.d5928

Higgins, J. P., & Green, S. (Eds.). (2008). *Cochrane handbook for systematic reviews of interventions*. Hoboken, JU: Wiley-Blackwell.

Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., . . . Bouter, L. M. (2007). Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology, 7*, 10. doi: 10.1186/1471-2288-7-10