Jonathan S. Comer, Ph.D. - Editor

Guidelines for Identifying Empirically Supported Treatments

Practical Recommendations for Clinical Researchers and Reviewers

David F. Tolin, Ph.D.
The Institute of Living
Yale University School of Medicine

Evan M. Forman, Ph.D.
Drexel University

E. David Klonsky
University of British Columbia

Dean McKay, Ph.D., ABPP
Fordham University

Brett Thombs, Ph.D.
Jewish General Hospital
McGill University

Recently, the Society of Clinical Psychology (SCP) updated its criteria for empirically supported treatments (ESTs). Whereas the original criteria (Chambless & Hollon, 1998; Chambless & Ollendick, 2001; Task Force on Promotion and Dissemination of Psychological Procedures, 1993) identified a psychological treatment as "well-established" when it was supported by at least two independently conducted, well-designed studies or a large series of well-designed and carefully controlled single case design experiments, and "probably efficacious" when it was supported by at least one well-designed study or a small series of single case design experiments, the new criteria (Tolin, McKay, Forman, Klonsky, & Thombs, in press) take advantage of the dramatic increase in published clinical trials over the past two decades, requiring the presence of systematic reviews of existing studies.

Based on the entire body of published research as synthesized in systematic reviews, treatments will now be assigned a recommendation level, derived from a modified version of the widely-used Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system (Atkins et al., 2004; Guyatt et al., 2008). The level of recommendation for a given psychological treatment may be Weak, Strong, or Very Strong. A Very Strong recommendation is made when there is high-quality evidence that the treatment produces a clinically meaningful effect on symptoms of the disorder being treated, as well as a clinically meaningful effect on functional outcomes, with significant improvement noted at immediate post-treatment and at a follow-up interval of not less than three months after treatment discontinuation, with relatively little risk of harm and reasonable resource use, and there is at least one well-conducted study that has demonstrated effectiveness of that treatment in non-research settings (e.g., settings that provide routine clinical care such as community mental health centers, inpatient or outpatient treatment facilities, health maintenance organizations, or private practices). A Strong recommendation requires the presence of moderate- to high-quality evidence that the treatment produces a clinically meaningful effect on symptoms of the disorder being treated, or on functional outcomes, again, with a clear positive balance in consideration of benefits versus possible harms and resource use. Evidence of external effectiveness of generalizability is not required for this level of recommendation. Weak recommendations are made when there is only low- or very low-quality evidence that the treatment produces a clinically meaningful effect on symptoms of the disorder being treated and/or functional outcomes, or when the evidence suggests that the effects of the treatment may not be clinically meaningful (though they may be statistically significant). When a given treatment does not merit one of the above recommendations, the Task Force will report on the reason(s) that the treatment was not recommended.

David F. Tolin

The aim of the present article is to guide researchers on how to produce and synthesize data in order

to obtain a recommendation for a psychological treatment according to the new EST criteria. We will work backwards through the process, beginning with the final step: the systematic review.

Developing systematic reviews that can be used to make EST recommendations

Systematic reviews will be evaluated by a Task Force, selected for breadth and depth of knowledge in psychological treatment and systematic reviews and absence of conflict of interest, operating under the SCP Committee on Science and Practice. The deliberations and findings of this Task Force will aim to be open and transparent at all times. The Task Force will evaluate published reviews as well as unpublished reviews which can be submitted by anyone, though it will not conduct its own reviews (that process will eventually be part of the American Psychological Association's Treatment Guidelines development process) (Hollon et al., 2014).

The Task Force will first evaluate the quality of a systematic review using an adaptation of the AMSTAR checklist (Shea, Bouter, et al., 2007; Shea, Grimshaw, et al., 2007; Shea et al., 2009). The aim of this checklist is to determine the degree to which a review's conclusions can be considered a reliable basis for clinical decision-making. The checklist is not used to generate a total score; accordingly, there is no cutoff at which a review is considered reliable; rather, the items on the checklist will be used to inform the group's decision of when a systematic review is of sufficient quality and reported sufficiently well. The checklist items give specific guidance for authors of systematic reviews. Specifically:

1. Use an 'a priori' design. Before the conduct of the review, define the research question and establish the study inclusion criteria. Ideally, systematic reviews will be registered with the PROSPERO international prospective register of systematic reviews.
2. Use duplicate study selection and data extraction. Have at least two independent data extractors, and develop a consensus procedure for disagreements.
3. Perform a comprehensive literature search. Search at least two electronic sources (e.g., MedLine, PsycInfo). In the report, describe the databases searched, as well as the publication years included in the search. List the search key words and/or MESH terms. Supplement the electronic search 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.

Evan M. Forman

4. State how you addressed publication status in study inclusion. For a comprehensive search, attempt to find unpublished reports as well as published ones. A search for unpublished reports could include searching Dissertation Abstracts International, posting requests for unpublished studies on relevant listservs, or other strategies. State whether or not any reports were excluded based on their publication status, language, or other factors.
5. Provide a list of included and excluded studies. A list of included and excluded studies should be provided. Many journals are unlikely to publish a list of studies that were not included; however, a list of excluded studies could be offered as online supplemental material or should at least be available upon request.
6. Describe the characteristics of the included studies. Create a table or other format in which you provide information about the participants, interventions, comparator and outcomes of each included intervention trial. Include sample information such as age, race, sex, relevant socioeconomic data, diagnosis, illness duration, illness severity, comorbidity, and concurrent treatments.
7. Assess the scientific quality of the included studies. Assessment and documentation of the quality of the reports is often overlooked in meta-analyses. In the next section, we will describe methods for evaluating risk of bias across relevant domains of clinical trial designs.

E. David Klonsky

8. Consider the scientific quality of the included studies when formulating conclusions. When conducting your analysis and developing conclusions, incorporate the methodological quality of the studies. When making recommendations, include an explicit statement about how the quality of the studies informs (for better or for worse) those recommendations.
9. Use appropriate methods to combine the findings of studies. When creating pooled results, use tests to ensure that it is appropriate to combine the studies. When significant heterogeneity among the studies is found, use a random effects model and/or make a logical argument about whether it is clinically appropriate to combine studies.
10. Assess the likelihood of publication bias. Include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).
11. State conflict of interest. This applies to the meta-analysis author as well as the authors of the included studies. For the meta-analysis author, acknowledge any sources of support or other potential conflicts of interest. For the included studies, indicate the degree to which conflicts of interest may constitute a risk of bias.
12. Calculate effect size estimates for both symptoms of the disorder and functional outcomes. A Very Strong recommendation is reserved for those treatments with a documented beneficial effect on both symptoms and functional outcomes.
13. Calculate effect size estimates at both post-treatment and at follow-up. For a Very Strong recommendation, clinically meaningful improvement must be documented not only at immediate post-treatment, but also at an interval of not less than three months after treatment discontinuation.
14. Identify studies that demonstrate effectiveness of the treatment in non-research settings. This study need not meet full inclusion criteria for the systematic review. However, in addition to the effect size estimates needed for the systematic review, A Very Strong recommendation also looks for at least one well-conducted study that suggests effectiveness of the treatment in settings that provide routine clinical care such as community mental health centers, inpatient or outpatient treatment facilities, health maintenance organizations, or private practices, not just in academic institutions.

Developing clinical trials that can be used for systematic reviews

Dean McKay

A systematic review is only as strong as the individual studies on which it is based. Therefore, it is important that clinical trial researchers produce high-quality studies that provide robust evidence for synthesis in meta-analyses. As noted above, it is incumbent on the authors of systematic reviews to evaluate the methodological quality of each of the included studies. The new SCP criteria (Tolin et al., in press) include an adaptation of the Cochrane Risk of Bias Tool (Higgins et al., 2011) for evaluating the quality of clinical trials. The items give specific guidance for authors of treatment outcome studies. Specifically:

1. Use an adequate sequence for allocating participants to treatments. There should be 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. There should be no non-random factors involved with assignment to groups.
2. Conceal allocation adequately. If clinical staff have knowledge about the groups to which the next patients recruited will be allocated, there is potential that this may influence who is recruited and when they are recruited, even if group assignments were initially made via randomization. Central allocation or sequentially numbered envelopes are both ways of concealing the allocation sequence, though central allocation, out of the hands of the research team, is the strongest method.
3. Keep study personnel and outcome assessors blind to treatment condition to the extent possible. In any clinical trial of psychological interventions, it is usually necessary to have some study personnel (e.g., clinicians, study coordinators) unblinded. However, at a minimum, outcome assessors should be unaware of participants' allocation, and measures should be used to assess whether the blind was broken.
4. When applicable, keep participants blind to

Brett Thombs

treatment condition. We recognize that in studies of psychological treatments, it is usually not possible to keep participants unaware of their treatment condition. In a study of treatment versus wait list, for example, participants are certainly aware of whether or not they are being treated. However, there may be some cases in which at least partial blinding is possible. Certain computerized treatments, for example, may permit randomization to conditions that are topographically similar, thus making it harder for participants to know whether they are receiving the active treatment (e.g., Amir, Beard, Burns, & Bomyea, 2009). In other cases, it might be appropriate to keep participants unaware of the study hypotheses, so that participants receiving two different treatments might not know which one is the target of the study. We recognize that this is a difficult aspect of psychological treatment research, and recommend that investigators consider different ways to prevent participants' knowledge of their treatment assignment from introducing systematic bias.

5. Use adequate strategies for handling incomplete outcome data. In an ideal clinical trial, there would be no missing outcome data. However, in reality, clinical trial results often have missing data due to attrition, skipped questions or questionnaires, equipment failure, and other factors. Primary trial outcomes should be evaluated on an intent to treat basis, which will typically involve the use of statistical imputation methods to take all of the available data into account. Clinical trials must be adequately powered to allow for such analyses; in many cases this will require substantially larger sample sizes than those that have been used in previously published trials. Completer analyses are not appropriate when there is missing data, and strategies such as last observation carried forward may yield misleading results.
6. Avoid selective outcome reporting. Before the study begins, identify the primary and secondary outcomes in a publicly-available study protocol or on a site such as www.clinicaltrials.gov. Ideally, a single primary outcome will be specified. In exceptional situations when more than one primary outcome is specified, appropriate statistical methods to account for multiple hypothesis tests must be described. The final paper should report on all outcomes specified in the pre-trial protocol with primary and secondary distinctions intact. In unanticipated situations, such as if data for the primary outcome cannot be obtained consistently, then changes in primary and secondary variables must be described.
7. Assess and document treatment fidelity. It is important to insure that the treatment was implemented as intended. Select therapists that have adequate qualifications and training to provide the study treatment. Use a publicly-available treatment manual so that others can replicate your findings. Monitor adherence to the treatment protocol in an ongoing fashion, using corrective measures such as additional training as needed.
8. Reviewing the adapted AMSTAR checklist for evaluating systematic reviews, clinical trial authors should also consider providing information that will feed into reviews that could generate a positive treatment recommendation. Specifically:
9. Describe the sample adequately. Provide information about your participants such as age, race, sex, relevant socioeconomic data, diagnosis, illness duration, illness severity, comorbidity, and concurrent and/or prior treatments.
10. Publish your results, whether or not your hypothesis was supported. Publication bias is a significant concern when reviewing the scientific literature, and it is important that the results of all clinical trials are disseminated. In the field of pharmaceutical research it is well documented that trials favorable to a sponsored product are more likely to be published than are trials not favorable to the sponsored product (Lexchin, Bero, Djulbegovic, & Clark, 2003; Lundh, Sismondo, Lexchin, Busuioc, & Bero, 2012). It is quite likely that the same phenomenon occurs in psychological treatment research as well. Registration of clinical trials (e.g., at www.clinicaltrials.gov) is increasingly emphasized to address this problem.
11. State conflict of interest. Acknowledge any sources of support or other potential conflicts of interest for the study.
12. Assess both symptoms of the disorder and

functional outcomes. The exclusive focus on symptom reduction risks ignoring other potentially important clinical outcomes, such as functional impairment (Dobson & Beshai, 2013). Although symptom reduction and improvements in functioning are significantly correlated, there can be a mismatch after treatment (see Vatne & Bjorkly, 2008, for review). Thus, it is possible that a treatment is highly effective at reducing specific target symptoms, and yet the patient fails to achieve desired clinical outcomes such as improved social or occupational functioning. We recommend that all clinical trials include at least one measure of work attendance or performance, school attendance or performance, social engagement, family functioning, or other functional measures.

13. Include follow-up assessments. Continue to assess study participants for at least three months after treatment discontinuation. In many cases, longer follow-up periods are desirable, such as in research involving addictive behaviors.
14. Conduct effectiveness research in addition to efficacy research. Effectiveness research focuses primarily on the generalizability of the treatment to more clinically representative situations. Criteria that could be considered include more diagnostically complex patients, effectiveness with non-randomized patients, effectiveness when used by non-academic practitioners, utility in open-ended, flexible practice, and outcomes in settings such as community mental health centers, inpatient or outpatient treatment facilities, health maintenance organizations, or private practices, not just in academic institutions.

Summary

As the quantity and quality of research on psychological treatments has increased, so too has the possibility and necessity of raising the bar for determining that a treatment is empirically supported. The new, more ambitious, criteria are described in detail elsewhere (Tolin et al., in press). The aim of the present article was to translate those criteria into tangible recommendations for investigators who wish to produce research that can be evaluated for EST recommendation.

The recommendation itself will be based on a transparent process using adapted AMSTAR criteria. Authors of systematic reviews, which can be submitted to the Task Force for review, should consider these criteria carefully. Specific recommendations include the use of 'a priori' designs, using duplicate study selection and data extraction, performing a comprehensive literature search and stating how publication status was addressed in study inclusion, providing a list of included and excluded studies, describing the characteristics of the included studies, assessing the scientific quality of the included studies and considering that quality when formulating conclusions, using appropriate methods to combine the findings of studies, assessing the likelihood of publication bias, stating conflict of interest, calculating effect size estimates for symptoms of the disorder and functional outcomes at both post-treatment and at follow-up, and identifying studies that demonstrate effectiveness of the treatment in non-research settings.

Similarly, clinical trial investigators can structure their research to more effectively and efficiently inform the systematic reviews. Meta-analysis authors are advised to evaluate clinical trials according to an adapted Cochrane Risk of Bias Tool. Clinical researchers are advised to consider the items on which the studies will be evaluated, including using an adequate sequence for allocating participants to treatments, concealing allocation adequately, keeping study personnel and outcome assessors (and participants, when appropriate and possible) blind to treatment condition, using adequate strategies for handling incomplete outcome data, avoiding selective outcome reporting, assessing treatment fidelity, providing adequate sample descriptions, publishing all trial results regardless of the outcome, stating conflict of interest, assessing symptoms of the disorder and functional outcomes at both post-treatment and at follow-up, and conducting both effectiveness and efficacy research.

We are the first to acknowledge that these recommendations set a very high bar for the quality of clinical trial reporting as well as the production of systematic literature reviews. However, we believe that the field has matured to the point where reaching these goals is quite possible. Furthermore, these recommendations are consistent with recommended procedures for developing guidelines for health care interventions, generally. Inevitably, some studies that were considered ESTs under the old criteria will not merit a recommendation under the new criteria, or there simply may not be enough research on a given treatment to conduct a systematic review at all. However, our hope is that like the previous criteria, the new criteria will stimulate a new generation of clinical research that provides clear evidence of the effects of psychological treatments, and that the dissemination of those findings will benefit consumers, practitioners, and policymakers. Ψ

References

Amir, N., Beard, C., Burns, M., & Bomyea, J. (2009). Attention modification program in individuals with generalized anxiety disorder. J Abnorm Psychol, 118(1), 28-33. doi: 2009-01738-016 [pii]10.1037/a0012589

Atkins, D., Eccles, M., Flottorp, S., Guyatt, G. H., Henry, D., Hill, S., . . . Group, G. W. (2004). Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. BMC Health Serv Res, 4(1), 38. doi: 10.1186/1472-6963-4-38

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. J Consult Clin Psychol, 66(1), 7-18. doi: 10.1037/0022-006X.66.1.7

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: controversies and evidence. Annu Rev Psychol, 52, 685-716. doi: 10.1146/annurev.psych.52.1.68552/1/685 [pii]

Dobson, K., & Beshai, S. (2013). The theory-practice gap in cognitive behavioral therapy: reflections and a modest proposal to bridge the gap. Behav Ther, 44(4), 559-567. doi: 10.1016/j.beth.2013.03.002

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., . . . Group, G. W. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. Bmj, 336(7650), 924-926. doi: 10.1136/bmj.39489.470347.AD

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. Bmj, 343, d5928. doi: 10.1136/bmj.d5928

Hollon, S. D., Arean, P. A., Craske, M. G., Crawford, K. A., Kivlahan, D. R., Magnavita, J. J., . . . Kurtzman, H. (2014). Development of clinical practice guidelines. Annu Rev Clin Psychol, 10, 213-241. doi: 10.1146/annurev-clinpsy-050212-185529

Lexchin, J., Bero, L. A., Djulbegovic, B., & Clark, O. (2003). Pharmaceutical industry sponsorship and research outcome and quality: systematic review. Bmj, 326(7400), 1167-1170. doi: 10.1136/bmj.326.7400.1167

Lundh, A., Sismondo, S., Lexchin, J., Busuioc, O. A., & Bero, L. (2012). Industry sponsorship and research outcome. Cochrane Database Syst Rev, 12, MR000033. doi: 10.1002/14651858.MR000033.pub2

Shea, B. J., Bouter, L. M., Peterson, J., Boers, M., Andersson, N., Ortiz, Z., . . . Grimshaw, J. M. (2007). External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS One, 2(12), e1350. doi: 10.1371/journal.pone.0001350

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 Med Res Methodol, 7, 10. doi: 10.1186/1471-2288-7-10

Shea, B. J., Hamel, C., Wells, G. A., Bouter, L. M., Kristjansson, E., Grimshaw, J., . . . Boers, M. (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol, 62(10), 1013-1020. doi: 10.1016/j.jclinepi.2008.10.009

Task Force on Promotion and Dissemination of Psychological Procedures. (1993). Training in and dissemination of empirically-validated psychological treatments: report and recommendation. The Clinical Psychologist, 48, 3-23.

Tolin, D. F., McKay, D., Forman, E. M., Klonsky, E. D., & Thombs, B. D. (in press). Empirically supported treatment: Recommendations for a new model. Clinical Psychology: Science & Practice.

Vatne, S., & Bjorkly, S. (2008). Empirical evidence for using subjective quality of life as an outcome variable in clinical studies A meta-analysis of correlates and predictors in persons with a major mental disorder living in the community. Clinical Psychology Review, 28(5), 869-889. doi: 10.1016/j.cpr.2008.01.001