Chapter 8
Step 3 of Evidence-Based Practice: Part 3—Meta-analysis and Systematic Reviews: Aggregating Research Results

Step 3 of the evidence-based practice (EBP) decision-making process is to critically evaluate the relevant research on your topic. In the previous two chapters, we have examined the role of research design and other methodological issues in evaluating practice research. These chapters explored how individual research studies are designed and reported in a rigorous manner. However, in many circumstances multiple research studies are available on a clinical topic. In this chapter we will explore how reports that aggregate the results of several separate studies on single topic are designed and analyzed.

Clinical social workers who search for research to use in EBP will often find compilations of several research studies on their topic of interest. Researchers call such compilations "meta-analyses" or "systematic reviews." Both can be very useful to the clinician seeking to evaluate research for use in practice. Both methods help the clinician appraise and compare the results of multiple studies on the chosen topic. It is important to understand how meta-analysis and systemic reviews are similar and different. It is also important to understand their strengths and limitations for evidence-based medicine (EBM) and for EBP.

Meta-analysis as a Method of Research Synthesis

Researchers first developed the method of *meta-analysis*, which introduced several important concepts and methods for aggregating research results. The steps of meta-analysis may seem quite familiar to clinical social workers and others learning EBP. Meta-analysis begins with (1) clearly formulating and stating the focal question, followed by (2) a defined and reproducible search of the literature. Next, researchers (3) evaluate the relevant literature using specifically stated quality criteria, leading to identification of the best research for analysis. Finally, studies are (4) compared statistically using specified procedures that vary to accommodate

J. W. Drisko, M. D. Grady, *Evidence-Based Practice in Clinical Social Work*,
Essential Clinical Social Work Series,
https://doi.org/10.1007/978-3-030-15224-6_8

different kinds of data. The process, in its outline, is parallel to some steps of the EBM/EBP practice decision-making process. The purpose of meta-analysis, however, is to aggregate study results and provide a general, summary conclusion.

Pratt, Rhine, Smith, Stuart, and Greenwood (1940) completed one of the earliest meta-analyses on the topic of extrasensory perception (ESP). They located and reviewed over 50 studies on ESP using similar methodologies. A full 61% of the studies endorsed ESP. Pratt and colleagues noted, however, that published reports with positive results had much more influence than did many more unpublished studies with negative results. This was important because, they believed, the unpublished studies might have been refused publication due to the negative results. In today's EBM/EBP context, early meta-analyses demonstrated how important it is to search thoroughly for all the relevant published and unpublished literature and to have a solid method for aggregating results.

In the 1970s and 1980s, researchers developed more sophisticated models of meta-analysis. Many of these pioneers addressed mental health topics. Smith, Glass, and Miller (1980) completed a meta-analysis entitled "The Benefits of Psychotherapy" that found similar, positive effects for several therapies for adults with depression or anxiety drawing on different theoretical premises and using different measures. A key innovation from Glass was a statistical method for aggregating studies that used different tests as outcome measures. These statistical techniques, called *meta-analysis*, are in wide use today in EBM/EBP.

It may be confusing that meta-analysis refers to both a research process with many steps as well as to statistical techniques. Meta-analysis set the stage for still more detailed systematic reviews of research. Today's systematic reviews employ the statistical techniques of meta-analysis in the aggregation of quantitative studies results. Yet articles may also be entitled "A meta-analysis of. . ." referring to either a systematic compilation of studies on a specific topic or only the use of meta-analytic statistical methods. Careful reading and critical thinking is required since authors may use the same terms quite differently.

Over time, wide variation in the quality of studies included in, or excluded from, meta-analyses proved a serious problem. Researchers might use very different standards to appraise research quality and very different methods to report how they had searched the literature. Issues of undisclosed biases, methodological flaws, and lack of reproducible results raised questions about the quality of many meta-analyses. Further, the methods used to locate studies were often poorly or inadequately described. These problems required the development of standards that would make compilations of research results fully reproducible by other researchers. Researchers call such reports *transparent* results. In EBM, the Cochrane Collaboration began in 1993. This international organization works to develop and promote standards for reviewing and synthesizing medical research results. Their standards added many details to the meta-analysis process but continued to use the meta-analysis statistical techniques.

The current, more refined approach to research synthesis is called the *systematic review*. The *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins & Green, 2011) defines a systematic review as an:

> attempt to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question. It uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing more reliable findings from which conclusions can be drawn and decisions made. (1.2.2)

A systematic review should have clearly stated objectives and eligibility/inclusion criteria for studies defined before the researcher begins the review. The methodology of the reviews should be reproducible, including details of the literature search strategy that produces the set of studies for review. Researchers should document the validity of included studies and make with extensive efforts to limit bias. The final presentation of the systematic review results should also be structured to provide details on the included studies (1.2.2).

The *Cochrane Handbook* further points out that many (but not all) systematic reviews contain meta-analyses. Meta-analysis here refers to only the statistical techniques for aggregating the results of several individual quantitative studies. "Meta-analyses can provide more precise estimates of the effects of health care than those derived from the individual studies included within a review... [by identifying] the consistency of evidence across studies, and the exploration of differences across studies" (Higgins & Green, 2011, 1.2.2).

In summary, meta-analyses introduced an approach to aggregating the results of research studies that was further refined and elaborated with the systematic review. Meta-analysis, as a set of statistical techniques, is now widely used as a component of a systematic review. Reports of systematic reviews can be a valuable resource for clinical social workers doing EBP. We will now explore the systematic review in detail.

Systematic Reviews

A systematic review is an aggregate summary of research on a single topic. Systematic reviews may focus on either quantitative or qualitative research studies, though the vast majority are quantitative. In the mental health literature, so-called systematic reviews may range from efforts based on an individual authors' work and views or a team effort following many well-defined rules to assure transparency and quality. At best, a systematic review is a special form of research synthesis, guided by an extensive set of rules and leading to both plain language and highly technical reports. These rules are intended to insure transparency of methods so that the review could be fully repeated, or replicated, by others. The rules establish procedures to locate, evaluate, and integrate the research results while limiting bias. Both the focus of the systematic review and its procedures must be defined before the review is begun. These practices are all steps to minimize bias and build transparency in the review.

The Cochrane Collaboration's *Handbook for Systematic Reviews of Interventions* (Higgins & Green, 2011) is the key resource for planning and implementing quantitative systematic reviews. The Campbell Collaboration also uses these standards

as the two organizations work cooperatively. The *Cochrane Handbook* is a book-length and very detailed document. It is updated from time to time to include new methods and improve the transparency of review procedures. (A major revision is planned for 2019.) The *Cochrane Handbook* is available for free review online at https://training.cochrane.org/handbook and also for purchase as a hardcover book.

Methods for systematic reviews of qualitative studies are more less standardized. Meta-ethnography, which pioneered the core methods of qualitative systematic reviews, was developed by Noblit and Hare (1988). Recent methods with a specific clinical focus are offered by Sandelowski and Barroso (2007), Dixon-Woods, Bonas et al. (2006), and Dixon-Woods, Cavers et al. (2006). The Cochrane Collaboration Methods Group (2017) offers an online library addressing methods of qualitative research synthesis. While qualitative systematic reviews have considerable merit, the focus of the remainder of this chapter will be on quantitative systematic reviews.

It is useful to understand what goes into a systematic review, even if most clinicians may not often review the full, technical versions of such reviews. More often, clinicians will use the much shorter plain language abstract that provides the key information in a condensed form. The format of even the plain language summaries follows that used for full systematic reviews. We remind readers that systematic reviews are not mentioned in the practice decision-making process of EBM/EBP. Systematic reviews are a process that groups of professionals have developed to summarize research results with rigor and consistency. Professional expertise and judgment must be used to decide if a systematic review offers more relevant and clinically useful information than do other sources of "the best available evidence" for practice decision-making.

Specifying the Systematic Review Topic

To start a systematic review, the topic it will cover must be clearly stated (Higgins & Green, 2011). This step is a guide to the researchers but also keeps the review focused. It also sets boundaries on what topics, and what kinds of research, will be studied from the outset. Its purpose is to keep systematic reviews focused and to limit alterations to the review to include material of interest to the authors but not quite on topic or not using high-quality methods.

Applications to register systematic reviews at both the Cochrane Collaboration and the Campbell Collaboration require a clear statement of the review focus. Scholars and researchers must apply to register a proposed systematic review before it is started. A team of peer reviewers evaluate each application. The peer reviewers look for a clear focus and a detailed research plan consistent with the *Cochrane Handbook* standards (Higgins & Green, 2011). Only proposals that meet the quality standards are accepted and registered. Clinicians may find some reviews in their preliminary stages listed online as "registered" or "in progress." Such reviews do not (yet) include any results. This can be frustrating as it does not help in practice

decision-making, but it shows such knowledge may be available in the near future. It also tells other researchers that the topic is taken.

Cochrane and Campbell systematic reviews are team efforts (Littell, Corcoran, & Pillai, 2008). Several people, ideally including at least one consumer, constitute the review team (Higgins & Green, 2011, section 2.3.4.1). Not only researchers but also policy-makers and clinicians may be included as well. This range of review team members should bring in some diversity of viewpoints about the purposes of the review and serve as some check on its final report. It also allows for more than a single researcher to assess each piece of research, again as an effort to limit bias.

Researchers undertaking a systematic review must also declare their "interests and commercial sponsorship" to set forth any potential conflicts of interest. Sponsorship of Cochrane reviews by any commercial interest is prohibited. Noncommercial sponsors of Cochrane systematic reviews, such as foundations offering research funding, must agree to have no influence over its process and final report content (Higgins & Green, 2011, 2.6).

Locating Research Studies for a Systematic Review

Once a topic is selected, the literature on the topic is next extensively reviewed. What literature is included in the review must be fully specified. It is common to see a listing of several electronic databases in quality systematic reviews. Each electronic database will cover many journals and articles. The point of detailing how the literature search was done in great detail is to make it replicable by other researchers. This means that other people could repeat the review process fully and should come up with the same results. It is a way of insuring quality and transparency in methods. Steps to insure quality and transparency are included in each part of the systematic review process.

Not only the databases that are included in the study are detailed; the search strategies used to locate relevant studies within each database are also set forth. These often include a combination of search terms or keywords that represent variations on the topic (i.e., depression, depressive, major depressive episode) but also types of research designs (i.e., experiments, RCTs, outcome studies). Further, the dates of the studies included in the review and the range of countries or languages searched are often stated. Again, the purpose of specifying the search strategy is to make the process both transparent to readers and replicable by other researchers.

Finally, systematic reviews include active efforts to locate unpublished research. There is a bias in journal publications that favors studies with positive results (Dickersin, 1990; Hopewell, Loudon, Clarke, Oxman, & Dickersin, 2009; McGauran et al., 2010; Sridharan & Greenland, 2009). "Positive results" are those that show a significant difference between a treatment and a control group or that one treatment is better than another. This means that research studies that do not show significant differences are often unpublished. Researchers call this the "file draw problem"

(Scargle, 2000). Research that is completed but unpublished may literally end up in a researcher's file draw, unknown to others. This can create a bias in favor of making it seem that a treatment is better than it would appear with the nonsignificant results also examined. Further, there have been cases where businesses and other groups that do not want negative results published have actively undermined the publication of negative results. One such case involves the active intervention of drug companies to limit publication of results that make their products look bad—even if they are not effective. Carey (2008) reports that makers of antidepressant drugs did not publish up to one-third of their research results but only those that were unflattering. Carey also notes that a similar incident had occurred previously in 2004.

Authors of systematic reviews seek out unpublished reports by contacting the author of published studies to ask if they have unpublished research on the topic. They also ask the published authors if they know of other researchers who might have unpublished studies. These efforts also uncover works in progress that have not been formally published, such as evaluations funded by state agencies. Occasionally the researcher is near completion of a study in progress. Such as yet unpublished results may be included in the systematic review as well, whether positive or negative. Efforts to locate unpublished results must also be specified in the systematic review methods.

Methods suggested to resolve problems that may distort the research available on a topic include publishing all studies and adding a second stage of peer review to the publication process (Carroll, Toumpakari, Johnson, & Betts, 2017). Until such methods are implemented, researchers instead seek to be fully transparent about the choices they make in including or excluding studies from a meta-analysis or a systematic review.

The entire process of the literature search can be presented in summary form using a Quorom flowchart (see Fig. 8.1). In the1990s, a group of physicians called the Quorom Group began to establish standards for reporting meta-analyses (Moher et al., 1999). Quorom is an acronym for the "*Qu*ality *o*f *R*eporting *o*f *M*eta-analyses." A Quorom flowchart summarizes the number of relevant items found at each stage of a literature search process. It starts with the total number of relevant articles or reports located and then identifies all the criteria for including or excluding reports in the review process. The flowchart identifies both why materials were included or excluded and how many reports were included or excluded. In the figure, we see that 9676 reports were located, but 8538 proved to be off topic. In the end, only 14 reports met all of the study's inclusion criteria and were not duplicates. A Quorom flowchart provides a quick and visually effective way to summarize the search process of a meta-analysis or a systematic review. They are beginning to become common in the social work literature (see, e.g., Drisko & Simmons, 2012; Litschge, Vaughn, & McCrea, 2010).

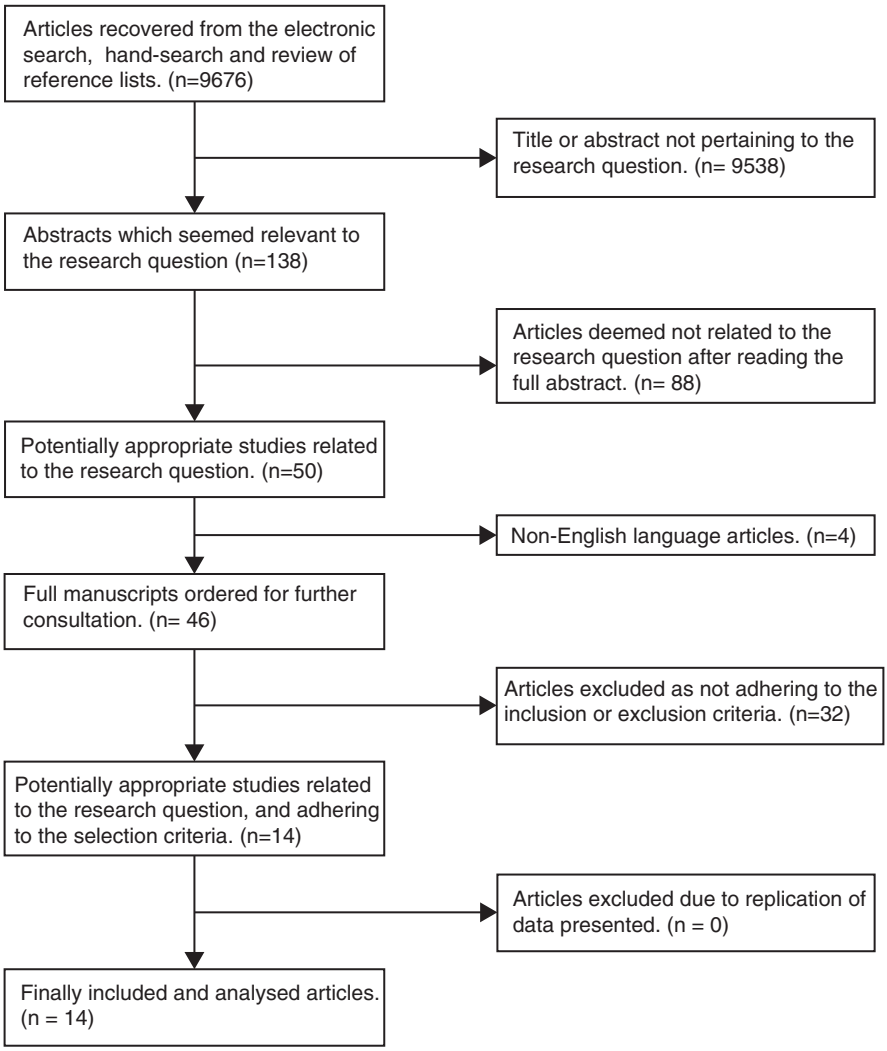

Fig. 8.1 A Quorom flow chart. (Retrieved from www.springerimages.com/Images/MedicineAndPublicHealth/5-10.1186_1471-2474-9-64-0 Source Chester et al. (2008)).

Evaluating the Methodological Quality of Research Reports in Systematic Reviews

The next activity in a systematic review is again parallel to the steps of the EBP process. Once research on the review topic is located, its quality must be evaluated. Most systematic reviews only include studies based on experimental or RCT

research designs. This is an acceptable choice for inclusion in a quantitative systematic review. It makes the rationale for including or excluding individual studies quite clear. On the other hand, it does not fully conform to the EBM/EBP standard of using "the best available" research evidence if such evidence is not based on experimental research. This may happen where a topic is not well conceptualized, such as chronic fatigue syndrome or reactive attachment disorder, or is not so common as to be extensively researched. The EBM/EBP practice decision-making process takes into account that not all topics are well studied by multiple RCTs. EBM/EBP process then points clinicians to use thoughtfully the results of research derived from other research designs, following the research design hierarchy. (Research design issues in EBP are examined in Chap. 6.)

While Cochrane systematic review standards clearly reflect a privileging of results from RCTs, there is also some ambiguity in the *Cochrane Handbook*. The *Cochrane Handbook* includes a brief section on including qualitative research results in systematic reviews (Higgins & Green, 2011, Chap. 20). This brief chapter states that qualitative research can address questions beyond those examined by quantitative research and may be helpful in explaining the results of RCTs. Here we find an instance where the hierarchy of research evidence used by both the Cochrane and Campbell Collaborations seems limited and other forms of research are explicitly acknowledged as valuable. Nonrandomized studies "may" be included in Cochrane reviews, but are to be "interpreted with caution" (Higgins & Green, 2011, Chap. 13). At the same time, qualitative studies are devalued as sources of evidence in other sections of the manual. Systematic reviews are social constructions and may have some inconsistencies despite a great deal of work by many people to make them internally consistent. Researchers differ in how they understand and prioritize research designs and methods. Systematic review standards may also apply a hierarchy of evidence that leads to results that are *less* inclusive of useful research knowledge than is called for by the core definitions of EBM and EBP.

The *Cochrane Handbook* emphasizes attention to reducing bias and ensuring clarity of methods in all reviews. This is a very worthy and important goal. Note carefully that much less attention is directed to the initial conceptualization of disorders, to the quality of measures, and to the definition of treatments. Experimental research has great strength in the attribution of cause-effect relationships but is only meaningful if the disorder of interest is quite well understood and the measures used to assess it are valid, complete, and reliable. Further, the populations included in the experimental research must be very similar to those of interest to clinicians. Attention to social diversity, and to socially structured oppressions, is not emphasized in the *Cochrane Handbook* criteria though it is heavily emphasized in social work.

Criteria for including or excluding studies in a systematic review may also be based on issues such as sample size, the clarity of the description of the treatment, the fidelity or care shown in insuring the treatment was delivered as described, and the kinds of statistical information provided. The various components of each study must be carefully assessed and coded as warranting inclusion in the systematic review or as warranting exclusion from it.

Coding Study Components and Results

All decisions regarding inclusion and exclusion from Cochrane and Campbell systematic reviews must be made by at least two reviewers. These reviewers work independently but then compare their results for consistency and quality control purposes. This ensures discussion where the reviewers are not fully in agreement regarding their views of a study's quality. How these differences were resolved, and the final inclusion/exclusion standards, must be stated in the full systematic review. At least two reviewers also make all later study quality appraisal decisions.

Summaries of the quality appraisal process may be presented in a Quorom flowchart in reports of systematic reviews. These flowcharts can illustrate succinctly the inclusion and exclusion criteria used in the review, as well as other criteria used in data extraction and assessment of study quality. As noted above, Quorom flowcharts specify the numbers of studies found in the literature search and the numbers included or excluded at each step in the review process. In most cases the number of research studies found on the topic is large and the final set of included studies much smaller.

Statistical Meta-analysis

As noted above, meta-analysis is a term used in two different but related ways. The first usage refers to a set of statistical techniques employed to combine the quantitative results of prior research. The yield of these statistics is usually a measure of *effect size*, a weighted average of the magnitude of difference between groups. This first, statistical usage will be our focus in this book. The other usage of meta-analysis now overlaps with the process of a systematic review. Indeed, meta-analyses were the precursors of today's systematic reviews. Meta-analyses, as publications, are research reports that locate, evaluate, and combine statistically the results of prior research. Systematic reviews expanded on the original methods of locating and evaluating research literature created for meta-analysis. Systematic reviews often use the statistical techniques of meta-analysis to combine the results of prior quantitative research. One key difference is that the procedures of Cochrane and Campbell systematic reviews are more fully detailed and better documented than is common in most meta-analysis reports. This increases the clarity and transparency of systematic reviews. Systematic reviews are also examined by methodologically sophisticated peer reviewers, which may not the case for some published meta-analyses.

The purpose of a meta-analysis statistic is to develop a common measure (also called a metric) to allow the combination of research results across several studies on a topic. Research results on a given topic often differ and may include discordant results. Meta-analysis provides a technique to combine results despite differences in measures. For example, in a meta-analysis of depression treatments, the goal is to combine the results of studies on people who have depression treated by different professionals, using different treatment models, and whose depression was assessed

by different measures. This might include people treated for depression with cognitive behavioral, interpersonal, short-term psychodynamic, and multisystemic therapies. The goal is to find out the overall effects of depression treatments and often to determine if the specific treatments have different results. Suppose it turns out that the available studies use different measures to assess depression levels. Some use the Hamilton Rating Scale for Depression and others use the Beck Depression Inventory. This may appear to be a major "apples and oranges" problem where the varied treatments and different measures do not seem to allow comparison. Solving this problem is the strength of meta-analysis. Meta-analysis allows the creation of a common measure of change or difference, using the same units despite any differences in the original scales used in each study. Thus, it allows for fair comparison of treatment outcomes despite initial differences in methods.

Statistics Used in Meta-analysis

The most widely used meta-analysis statistic is Cohen's *d* (Ellis, 2010), also called a standardized mean difference (SMD). Cohen's *d* is a measure of the standard mean difference across two groups. This statistic shows the magnitude of change in an experimental comparison. Researchers calculate it by subtracting the mean score for the control group from the mean score of the treated group and then dividing the result by the pooled standard deviation of both groups. Cohen's *d* provides a measure of the magnitude of the difference between the treatment and control groups, taking into account the variation of scores within both groups.

Cohen's *d* scores range from 0.0 to about 2.0. Cohen (1988) established rough benchmarks to help interpret these scores in plain language. Cohen's *d* values between 0.00 and 0.49 are "small" effects. Values from 0.50 to 0.79 are "moderate," and values larger than 0.80 are "large" effects. This means that the magnitude of the difference between two treatments can be called "small" if Cohen's *d* is 0.48 or lower and "large" if the Cohen's *d* value is 0.80 or larger. Lenth (2008) cautions that all effects size statistics should be interpreted in context and with attention to operational definitions.

The purpose of Cohen's *d* is to assess the magnitude of differences between groups. This is not the same as establishing the statistical significance (usually reported with a p = value). However, it is *usually* the case that moderate and strong effect sizes are found where significant differences between groups are found using probability tests. Effect size measures complement probability statistics.

Cohen's *d* is best when used with large samples. With smaller samples (roughly 20 or fewer participants in each treatment), the Hedges' *g* statistic is usually a better choice. This is because in a small sample one or two "outliers," or extreme scores, can have a large impact on the value of the *d* statistic. In addition, Hedges' *g* employs a different calculation of the pooled standard deviation than does Cohen's *d*. It includes a correction factor when used to make population level estimates. Thus, Hedges' *g* yields more conservative estimates of the magnitude of group differences

for small samples. Some critiques of mental health meta-analyses target the use of Cohen's *d* with small samples as a methodological concern (Thombs & Jewett, 2009). Hedges' *g* does not make any statement about whether or not the difference found is likely to reflect that in the larger population, as probability statistics do. The Hedges' *g* statistic is very rarely covered in social work research and statistics texts.

SMD statistics are usually coupled with a *95% confidence interval*. A 95% confidence interval is an estimated range likely to include a value (technically a population parameter) of interest. This value is generally unknown, so the CI is a statistical estimate of its likely value in the whole population. Two numbers define a 95% CI, a lower and an upper value around which it is probable with 95% confidence that the true population value lies (O'Brien & Yi, 2016). In other words, if a standardized mean difference (or SMD; such as Cohen's *d* or Hedge's *g*) value is found within the 95% CI, it is likely that, in the population from which the sample was drawn, the results will be similar to the true population value. Thus, the result has practical implications for the entire population, not just the specific sample that was studied. In a systematic review report, one might find information such as, for symptom reduction, individual CBT and EMDR were more effective than was a control group (SMD −1.79; 95% CI −2.12 to −1.16). Because the SMD value is within the range defined by the 95% CI, it is likely that the unknown population value is likely to be like the study results. This supports confidence for its use in practice.

Other measures of effect size include correlations and odds ratios. Correlations are used as measures of effect sizes in observational studies. Correlations statistics can serve as their own measures of effect size, with a range from −1.0 (a perfect negative correlation, through 0.0 (no correlation) to +1.00 (a perfect positive correlation). Cohen (1988) states that a correlation of 0.10 may be interpreted as "small," a correlation of 0.25 or larger as "medium," and a correlation of 0.40 or higher as "large."

An *odds ratio* (OR) is a summary measure of the association between a treatment or "exposure" and an outcome where both variables are binary (Szumilas, 2010). It summarizes the odds or chance that an outcome will occur given a particular exposure in contrast to the odds of the same outcome occurring without the exposure. Odds ratios are most commonly used in nonexperimental case-control studies. Odds ratios are interpreted as an OR of exactly 1 indicating that exposure does not affect odds at outcome; an OR of less than 1 indicating that exposure is associated with lower or reduced odds at outcome (Deng, 2012); and OR > 1 indicating that exposure is associated with greater or increased odds of risk at outcome. OR results do not include statistical significance, nor do they demonstrate cause and effect. A 95% confidence interval (CI) is typically coupled with an OR to provide an estimate of the precision of an OR result. A large CI indicates a low level of precision; in contrast a small CI indicates a higher precision of the OR. An example of OR is found in a Cochrane Collaboration systematic review of treatments for post-traumatic stress disorder [PTSD] in children and adolescents. Gillies, Taylor, Gray, O'Brien, and D'Abrew (2012) found that "across all psychological therapies [studied], improvement was significantly better (three studies, n = 80, OR 4.21, 95% CI 1.12 to 15.85) … compared to a control group" (Abstract, Main results). In other words, having

treatment (the exposure) was associated with more than four times greater odds of improvement in PTSD symptoms than was no treatment. However, the CI, ranging from 1.12 to 15.85, suggests a low level of precision for these results as the CI is a large range (14.73) compared to the OR of 4.21. The authors go on to state that "the findings of this review are limited by the potential for methodological biases, and the small number and generally small size of identified studies" (Author's conclusions). This may, in part, account for the low level of precision of the CI in this study.

Relative risk (RR), also known as the risk ratio, is another statistic that measures the risk of an event relative to an independent variable (Chittaranjan, 2015). Relative risk is a ratio of the probability of the event occurring in a treated or "exposed" group versus an untreated or "nonexposed" group (Sistrom & Garvan, 2004). The frequency, or count, of how often the event occurs in the treated group is divided by the frequency of the event observed in the untreated group. For example, the number of smokers who develop lung cancer might be compared over a 20-year period to the number of nonsmokers who develop the same illness. A relative risk ratio value of 1.0 indicates no difference in risk between the two groups. A value of less than 1.0 indicates the outcome event is less likely to occur in the treated group than the control/comparison group. A value of more than 1.0 indicates the outcome event is more likely to occur in the treated group than in the control/comparison group. OR and RR are epidemiological statistics that are rarely covered in social work research and statistics texts. They are, however, quite commonly found in systematic reviews.

Risk statistics may be reported along with "Number Needed to Treat" (NNT) statistics. NNT is a measure used in epidemiology to assess the effectiveness of a treatment (Christensen & Kristiansen, 2006; Mendes, Alves, & Batel-Marques, 2017). *NNT is the number of people who need to be treated to prevent one additional negative outcome*. In other words, how many people would need to be treated for one person to benefit more than was observed in the control/comparison group. A NNT value of 1 is ideal since it means everyone who is treated benefits from the treatment while no one benefits in the comparison group. Low NNT values indicate that the treatment is more effective compared to the comparison group. The higher the NNT value, the fewer people benefit compared to the comparison group in the study. McQuay and Moore (1997) note that NNT values of 2 or 3 are rare but indicate that the treatment is highly effective.

Researchers use many statistics with meta-analyses depending on the nature of the data relevant to the clinical question. A more complete overview of these statistics may be found in Littell et al. (2008). The European Patients' Academy (2015) provides an online tutorial for patients and families at www.eupati.eu/clinical-development-and-trials/statistics-clinical-trials-key-concepts/.

Meta-analytic statistical analyses are not possible when only one or two studies are found on a topic. This is why meta-analysis may not be appropriate for all systematic reviews. Such a small number of studies does not allow for appropriate use of the meta-analysis statistics. Of course, meta-analysis statistics are not used in qualitative systematic reviews.

Meta-analysis is a very helpful part of a quantitative systematic review. The procedures of the systematic review structures the identification of high-quality studies

that meet the stated inclusion criteria for the review. A meta-analysis provides a useful summary statistical measure of the magnitude of differences between the tested groups. It also provides a general way to articulate the statistical differences in plain language.

Integrating Results and Identifying Limitations and Cautions

Meta-analysis is one very useful statistical method for integrating the results of several studies on a topic. Written Cochrane and Campbell Collaboration systematic reviews generally include meta-analysis statistical results. In addition, most systematic reviews also point out the limitations of the available research and some cautions about its application. For example, a systematic review might note that while two treatments are effective, the outcome for one treatment was measured by rates of rehospitalization and the other was measured by a standardized test of symptoms. Clearly helping a client stay out of the hospital is a general good, but it may not be the main concern of the client or the client's family. The client may be more directly concerned about reducing symptoms. In such a situation, the systematic review may help the clinician recommend to the client the treatment that demonstrated symptomatic improvement, while noting an alternative treatment was also effective but that it was measured by rates of rehospitalization. Here the details of the review can be useful in helping the client understand the differences in evidence that may matter to them in making informed treatment choices.

Systematic review summaries may also point out the relative limitations of the available research they summarize. For example, review authors may state that sample sizes for all of the included treatments were small. This might suggest that while the treatments were effective, it is not yet unclear if this will generalize to other clients and settings. Cochrane and Campbell systematic reviews also point out potential biases in the available research. One key area is to identify when the researchers who study a treatment are also the originators of the treatment model under investigation. Considerable research suggests that when the developer of a treatment model tests the effectiveness of their model they may, unconsciously or consciously, introduce attribution biases in favor of their model. It is helpful for readers to keep in mind what kinds of biases might influence studies done by researchers who might favor one model over another. Economic influences, such as sources of funding, may also shape research results in a nonscientific, biased manner.

An Example of a Cochrane Collaboration Systematic Review

One disorder of considerable interest to clinical social workers is post-traumatic stress disorder [PTSD]. Adult clients present with PTSD based on a wide range of traumatic and overwhelming experiences. Their stories may also generate painful reactions in clinicians and others. PTSD is linked with several other disorders,

notably including personality disorders, and may be difficult to differentially diagnose. Its presentation may be delayed following the trauma, and it may present with dissociation.

The DSM-5 criteria (American Psychiatric Association, 2013) center on exposure to risk of death, serious injury, or sexual violence. Exposure may be direct, through witnessing, or indirect through learning of a loved one's exposure to trauma or through indirect exposure to traumatic details. These exposures lead to persistent re-experiencing of the trauma through memories, nightmares, flashbacks, or emotional and/or physical reactivity after reminders of the event. People avoid trauma-related stimuli, thoughts, feelings, or reminders. Symptoms must last more than 1 month and include worsening of trauma-related arousal and reactivity that cause significant distress or impairment of functioning. Symptoms must not be due to medications, substance use, or other illnesses.

A search of the Cochrane Library reveals 26 systemic reviews related to PTSD. Medications, psychotherapy, sports and games, and diaries are treatments included in these systematic reviews. Target clinical populations are adults, children, and adolescents, victims of torture, law enforcement officers, women following childbirth, as well as people with critical illnesses and traumatic injuries.

Some preventive efforts are also included in the Cochrane systematic reviews. A systemic review by Rose, Bisson, Churchill, and Wessely (2002) found single-session debriefing to be ineffective. Amos, Stein, and Ipser (2014) report in their systematic review of the effectiveness of mediation for preventing PTSD that:

> There is moderate quality evidence for the efficacy of hydrocortisone for the prevention of PTSD development in adults. We found no evidence to support the efficacy of propranolol, escitalopram, temazepam and gabapentin in preventing PTSD onset. The findings, however, are based on a few small studies with multiple limitations. Further research is necessary in order to determine the efficacy ofpharmacotherapy in preventing PTSD and to identify potential moderators of treatment effect. (Author's conclusions)

Cochrane Systematic Reviews: Organization and Audiences

Cochrane Library reviews located online typically provide only a summary or abstract of available information. This may be sufficient for many information needs. However, pdfs of the more complete *standard* version and the entire *full* versions are also available, though at cost for US users. The standard version provides more detail than does the abstract alone but excludes complete data and analyses as well as appendices. The full version is the complete report. Only the full version reveals all the choices, and rationales for them, used in completing the SR. Full versions of SR reports can be over 200 pages in length.

Psychological Therapies for Chronic Post-traumatic Stress Disorder (PTSD) in Adults

Bisson, Roberts, Andrew, Cooper, and Lewis (2013) completed a systematic review [SR] of psychological therapies for adults who have chronic PTSD registered with the Cochrane Collaboration. Their SR is available in English, Spanish, and Portuguese languages. Information is included for clinicians as well as for clients and the public. Both detailed professional and plain language summaries are included in the SR report. The plain language summary of this SR, intended for both clinicians and for clients, is available online at https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD003388.pub4/full#CD003388-abs-0004.

The Plain Language Summary of a Systematic Review

Bisson et al.'s (2013) plain language summary begins with a section on *background information* defining PTSD and continuing to explain some previously researched therapies for it:

> Post-traumatic stress disorder (PTSD) can occur following a traumatic event. It is characterised by symptoms of re-experiencing the trauma (in the form of nightmares, flashbacks and distressing thoughts), avoiding reminders of the traumatic event, negative alterations in thoughts and mood, and symptoms of hyper-arousal (feeling on edge, being easily startled, feeling angry, having difficulties sleeping, and problems concentrating).
>
> Previous reviews have supported the use of individual trauma-focused cognitive behavioural therapy (TFCBT) and eye movement desensitisation and reprocessing (EMDR) in the treatment of PTSD… (Plain language summary, Background)

The next section of the plain language summary describes the *Study characteristics* of the studies included in the SR: "This review draws together up-to-date evidence from 70 studies including a total of 4761 people" (Study characteristics). Here we learn how many studies and participants were included in the SR, but do not learn why other studies were included or excluded by the authors.

Key findings make up the next section of the SR's plain language summary. These findings link to the treatments defined previously in the background section of the SR:

> There is continued support for the efficacy of individual TFCBT, EMDR, non-TFCBT and group TFCBT in the treatment of chronic PTSD in adults. Other non-trauma-focused psychological therapies did not reduce PTSD symptoms as significantly. There was evidence that individual TFCBT, EMDR and non-TFCBT are equally effective immediately post-treatment in the treatment of PTSD. There was some evidence that TFCBT and EMDR are superior to non-TFCBT between one to four months following treatment, and also that individual TFCBT, EMDR and non-TFCBT are more effective than other therapies. No specific conflicts of interest were identified [among the researchers competing the original studies]. (Key findings)

The last section of the plain language summary offers a *quality appraisal* of the best available research evidence as determined by the SR authors:

> Although we included a substantial number of studies in this review, each only included small numbers of people and some were poorly designed. We assessed the overall quality of the studies as very low and so the findings of this review should be interpreted with caution. There is insufficient evidence to show whether or not psychological therapy is harmful.

Overall, the plain language summary of the Bisson et al. (2013) SR indicates that psychological therapies are more effective than no treatment for persons who have chronic PTSD, with trauma-focused therapy and EMDR supported by the best available research evidence, though the evidence is of low quality and based on few studies. No harms from undertaking these therapies were reported. In addition, the *Author's conclusions* state that "non-TFCBT is effective in the short term" and "there is more limited evidence that some other non-trauma-focused psychological therapies may be effective" (Bisson et al., 2013, Author's conclusions).

Plain language summaries offer a clear and succinct way to access the best available evidence under Cochrane Collaboration standards. Considerable information is summarized, though even the plain language summary may be confusing to many readers. It is a very useful resource and starting point for professionals doing EBP in practice. However, much greater detail is also available in a full SR report. Plain language summaries may be very useful in Step 4 of the EBP process in which the clinician must inform the client about the best available research evidence in treatment planning.

The Systematic Review Abstract

Like the plain language summary, the SR abstract opens with background information about the disorder under study and the objectives of the SR. A key difference is that the *Search methods* employed are described in much greater detail. This is to better guide the clinician and to show which databases and publications were examined. Efforts to locate unpublished studies and studies in progress are also stated. (Complete details are only found in the *Full* SR report however.)

Bisson et al. (2013) state:

> For this update, we searched the Cochrane Depression, Anxiety and Neurosis Group's Specialised Register (CCDANCTR-Studies and CCDANCTR-References) all years to 12th April 2013. This register contains relevant randomised controlled trials from: The Cochrane Library (all years), MEDLINE (1950 to date), EMBASE (1974 to date), and PsycINFO (1967 to date). In addition, we hand searched the Journal of Traumatic Stress, contacted experts in the field, searched bibliographies of included studies and performed citation searches of identified articles. (Search methods)

Specific search terms are not detailed, but the approach of the SR team and their database search parameters are clearly stated. Additional searches for the gray literature were also undertaken using several techniques.

Next, a section describes the *Selection criteria* or more precisely the inclusion and exclusion criteria for this SR:

> Randomised controlled trials of individual trauma-focused cognitive behavioural therapy (TFCBT), eye movement desensitisation and reprocessing (EMDR), non-trauma-focused CBT (non-TFCBT), other therapies (supportive therapy, non-directive counselling, psychodynamic therapy, and present-centred therapy), group TFCBT, or group non-TFCBT, compared to one another or to a waitlist or usual care group for the treatment of chronic PTSD. The primary outcome measure was the severity of clinician-rated traumatic-stress symptoms.

Note that consistent with the Oxford University Hierarchy of Research Evidence (, 2016), only experimental studies or RCTs were included in this SR. RCTs allow cause-effect determination unlike other research designs. However, other potentially informative studies using nonexperimental research designs were excluded. Note too that a clear definition of PTSD using DSM and ICD standards is assumed, though studies may use different versions of PTSD diagnoses over time with different criteria (such as the older DSM-III or DSM-IV criteria). Further, the kinds of trauma leading to PTSD in these adults are not specified. Many kinds of trauma appear to be included, without consideration of specific differences in populations needs across victims of sexual violence versus combat veterans.

Next, the *Main results* section offers much more detail on findings than does the plain language summary. Beyond the number of studies and total number of participants, we now find details about specific comparisons among treatment, including statistics supporting the authors' conclusions:

> We include 70 studies involving a total of 4761 participants in the review. The first primary outcome for this review was reduction in the severity of PTSD symptoms, using a standardised measure rated by a clinician. For this outcome, individual TFCBT and EMDR were more effective than waitlist/usual care (standardised mean difference (SMD) -1.62; 95% CI -2.03 to -1.21; 28 studies; n = 1256 and SMD -1.17; 95% CI -2.04 to -0.30; 6 studies; n = 183 respectively). There was no statistically significant difference between individual TFCBT, EMDR and Stress Management (SM) immediately post-treatment although there was some evidence that individual TFCBT and EMDR were superior to non-TFCBT at follow-up, and that individual TFCBT, EMDR and non-TFCBT were more effective than other therapies.
>
> Non-TFCBT was more effective than waitlist/usual care and other therapies. Other therapies were superior to waitlist/usual care control as was group TFCBT. There was some evidence of greater drop-out (the second primary outcome for this review) in active treatment groups. Many of the studies were rated as being at 'high' or 'unclear' risk of bias in multiple domains, and there was considerable unexplained heterogeneity; in addition, we assessed the quality of the evidence for each comparison as very low. As such, the findings of this review should be interpreted with caution. (Main results)

Here we see that clinician ratings of symptom severity using a standardized measure were the data source for the outcome or dependent variable. Individual TFCBT and EMDR were more effective than either "usual care" or a wait list condition. The effect size or SMD [standardized mean difference] for TFCBT was −1.62, a large reduction from initial symptom severity at the end of treatment. The reported 95% confidence interval ranges from −2.03 to −1.21. (The negative signs indicate reduction in PTSD symptoms, the desired goal of treatment.) Since the SMD of −1.62

falls within the 95% CI range, it is likely that these results will prove applicable to the larger population of adults with PTSD. Similarly, the SMD for EMDR was −1.17 which falls within the 95% CI range of −2.04 to −0.30.

Note that there was no statistically significant difference among TFCBT, EMDR, and stress management at the end of treatment, though TFCBT and EMDR appeared with "some evidence" to be superior to stress management at later follow-up. Further, individual TFCBT, EMDR, and non-TFCBT were more effective than were "other therapies" though the "other therapies" proved superior to wait list/usual care control along with group TFCBT. Complete supporting details for these conclusions can be found in the full version SR report.

The greater detail provided in the abstract of this SR can help clinicians answer client questions about specific treatments not discussed in the plain language summary. This is especially important if the treatments with the strongest research support are not acceptable to the client due to the client's values and preferences or are not accessible due to financial limitations or lack of appropriately trained clinicians in a geographical area.

The last section of the SR abstract addresses the *Authors' conclusions*. Here the SR authors describe the quality of the studies they located and included in the SR. Not all studies are of high quality, though they are included in the SR as "the best available evidence" meeting their stated inclusion criteria. Bisson et al. (2013) state:

> The evidence for each of the comparisons made in this review was assessed as very low quality. This evidence showed that individual TFCBT and EMDR did better than waitlist/usual care in reducing clinician-assessed PTSD symptoms. There was evidence that individual TFCBT, EMDR and non-TFCBT are equally effective immediately post-treatment in the treatment of PTSD. There was some evidence that TFCBT and EMDR are superior to non-TFCBT between one to four months following treatment, and also that individual TFCBT, EMDR and non-TFCBT are more effective than other therapies. There was evidence of greater drop-out in active treatment groups. Although a substantial number of studies were included in the review, the conclusions are compromised by methodological issues evident in some. Sample sizes were small, and it is apparent that many of the studies were underpowered. There were limited follow-up data, which compromises conclusions regarding the long-term effects of psychological treatment. (Author's conclusions)

Here we see that though some treatments prove more effective than wait list or usual care, these conclusions must be tempered by the "very low quality" of the available outcome studies for chronic adult PTSD. Unstated methodological concerns (which would be detailed in the full version of this SR) also were found in some included studies. One concern was the lack of statistical power (the ability of a statistic to demonstrate *any* difference). The lack of many follow-up studies also limited evaluation of how lasting these changes were after the end of treatment.

The abstract summary of this SR provides more detail than does the plain language summary version. With greater detail comes a requirement for the reader to have the knowledge and skill to interpret issues of research design and statistics. This requirement increases further when one reads and interprets the full version of the SR.

Note carefully that the treatments under consideration are broadly defined, but shared or common components are not mentioned. Summary definitions of the various treatments are found in the full version SR (see pp. 6–7). The full version also notes that there was no restriction on including studies that treated persons with comorbid conditions as well as PTSD, so long as PTSD was the primary diagnosis (p. 7).

SRs help clinicians locate relevant research studies and critically evaluate research results. However, critical thinking is always required, as is careful attention to how well studies and SRs include persons who are like the client you are treating. Most SRs do not provide details on socially constructed oppressions and other factors that may impact treatment feasibility and effectiveness.

Evaluating Systematic Reviews

We have noted that systematic reviews and meta-analysis are terms that may be applied to reports of varying methods and overall quality. Working groups have made efforts to establish ways to assess the methodological quality of systematic reviews. One such effort is AMSTAR, developed by an international group of epidemiologists and public health specialists (Shea et al., 2007). AMSTAR is an acronym for *A*ssessment of *M*ultiple *S*ys*T*em*A*tic *R*eviews. AMSTAR is an 11-item rating scale for assessing systemic reviews, including meta-analytic content. The full checklist is available free online at https://amstar.ca/Amstar_Checklist.php. The checklist summarizes several quality issues necessary for a strong systematic review. These include having an a priori review design (stating the review question at the start) and clear report selection criteria. Other criteria address standards for "scientific quality" assessment of each included study and for appraising the quality of meta-analyses used to synthesize multiple studies. Finally, the use of a range of efforts to identify and reduce bias in the systematic review is assessed. Cochrane and Campbell systematic reviews generally conform to the AMSTAR criteria very well.

The AMSTAR checklist provides a useful way for clinicians to frame their assessments of systematic reviews. Its main limitation is that the 11 items are quite general. For example, specific elements of "scientific quality" assessment are not provided but must be undertaken by the user. This leaves many technical issues to the knowledge and skills of the user. On the other hand, the AMSTAR checklist provides a very helpful framework for assessing the overall quality of a systemic review.

Summary

This chapter has reviewed systematic reviews as used in clinical research and in Step 3 of EBM/EBP process. It also introduced the Cochrane and Campbell Collaboration systematic reviews process and requirements. One part of this

process is the statistical technique known as meta-analysis. Meta-analysis provides a way to combine the quantitative results of several different research studies. Clinical social workers are reminded that not all reports claiming to be systematic reviews meet the careful standards of the *Cochrane Handbook* (Higgins & Green, 2011). Similarly, meta-analytic reports studies may vary in scope and quality.

The reports of Cochrane and Campbell systematic reviews include both a plain language summary of key results and a much longer technical report giving much more detail about each study and the review process. In the United States, only the abstracts of Cochrane Collaboration systematic reviews are available free online. The full technical reports, however, are only available on a subscription basis. In most other countries, the full versions of the Cochrane systematic reviews are available free. Even abstracts of systematic reviews may include considerable statistical detail. There are both print and online resources that can help clinicians interpret these details as needed.

While the EBM/EBP practice decision-making model makes no mention of systematic reviews or meta-analyses, the standards of Cochrane and Campbell reviews make them a key source for clinicians. Still, clinical social workers must apply their professional expertise and critical skills in determining just what constitutes "the best available research evidence" for their unique practice needs.

Appraising the quality of research reports and systematic reviews requires skills quite different from the core skills of clinical practice or the skills used in locating practice research. It can be a complex, multifaceted process. Yet bringing the results of high-quality research into practice decision-making should improve outcomes and reduce harm. It is an important part of EBM and EBP.

The next step in the EBP practice decision-making process is to collaboratively discuss the best available research with the client. Step 4 brings "the best available evidence" back to the client to inform their part of treatment planning and to allow for the client's questions and concerns to be addressed. This process is the focus of the next chapter.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.) (DSM-5). Washington, DC: Author.

Amos, T., Stein, D. J., & Ipser, J. (2014). Pharmacological interventions for preventing post-traumatic stress disorder (PTSD). *Cochrane Database of Systematic Reviews, 2014*(7), CD006239. https://doi.org/10.1002/14651858.CD006239.pub2

Bisson, J., Roberts, N., Andrew, M., Cooper, R., & Lewis, C. (2013). Psychological therapies for chronic post-traumatic stress disorder (PTSD) in adults. *Cochrane Database of Systematic Reviews, 2013*(12), CD003388. https://doi.org/10.1002/14651858.CD003388.pub4

Carey, B. (2008, January 17). Antidepressant studies unpublished. *The New York Times.* Retrieved from https://www.nytimes.com/2008/01/17/health/17depress.html

Carroll, H., Toumpakari, Z., Johnson, L., & Betts, J. (2017). The perceived feasibility of methods to reduce publication bias. *PLoS One, 12*(10), e0186472. https://doi.org/10.1371/journal.pone.0186472

Chester, R., Smith, T., Sweeting, D., Dixon, J., Wood, S., & Song, F. (2008). The relative timing of VMO and VL in the aetiology of anterior knee pain: a systematic review and meta-analysis. *British Medical College Musculoskeletal Disorders, 9*(1), 64. https://doi.org/10.1186/1471-2474-9-64

Chittaranjan, A. (2015). Understanding relative risk, odds ratio, and related terms: As simple as it can get. *The Journal of Clinical Psychiatry, 76*(7), e857–ee86.

Christensen, P., & Kristiansen, I. (2006). Number-needed-to-treat (NNT): Needs treatment with care. *Basic & Clinical Pharmacology & Toxicology, 99*(1), 12–16.

Cochrane Methods (Group): Qualitative and implementation. (2017). *Core library of qualitative synthesis methodology* [online bibliography]. Retrieved from: http://methods.cochrane.org/qi/core-library-qualitative-synthesis-methodology

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Routledge Academic.

Deng, C.Q. (2012, February 12). *How to interpret odds ratios that are smaller than 1?* Retrieved from https://onbiostatistics.blogspot.com/2012/02/how-to-interpret-odds-ratios-that-are.html

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association, 263*, 1385–1389.

Dixon-Woods, M., Bonas, S., Booth, A., Jones, D., Miller, T., Sutton, J., et al. (2006). How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research, 6*(1), 27–44.

Dixon-Woods, M., Cavers, D., Agarwal, S., Annandale, E., Arthur, A., Harvey, J., et al. (2006). Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Medical Research Methodology, 6*. (electronic journal). Retrieved from http://www.biomedcentral.com/1471-2288/6/3

Drisko, J., & Simmons, B. (2012). The evidence base for psychodynamic psychotherapy. *Smith College Studies in Social Work, 82*(4), 374–400.

Ellis, P. (2010). *The essential guide to effect sizes: An introduction to statistical power, meta-analysis and the interpretation of research results*. New York: Cambridge University Press.

European Patients' Academy. (2015). *Statistics in clinical trials: Key concepts*. Retrieved from https://www.eupati.eu/clinical-development-and-trials/statistics-clinical-trials-key-concepts/

Gillies, D., Taylor, F., Gray, C., O'Brien, L., & D'Abrew, N. (2012). Psychological therapies for the treatment of post-traumatic stress disorder in children and adolescents. *Cochrane Database of Systematic Reviews, 2012*(12), CD006726. https://doi.org/10.1002/14651858.CD006726.pub2

Higgins J., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions,* version 5.1. Available from https://training.cochrane.org/handbook

Hopewell, S., Loudon, K., Clarke, M. J., Oxman, A., & Dickersin, K. (2009). Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database of Systematic Reviews, 2009*(1), MR000006. https://doi.org/10.1002/14651858.MR000006.pub3

Lenth, R. (2008). *Java applets for power and sample size*. Retrieved from http://www.stat.uiowa.edu/~rlenth/Power/

Litschge, C., Vaughn, M., & McCrea, C. (2010). The empirical status of treatments for children and youth with conduct problems: An overview of meta-analytic studies. *Research on Social Work Practice, 20*(1), 21–35.

Littell, J., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. New York: Oxford University Press.

McGauran, N., Wiesler, B., Kreis, J., Schüler, Y.-B., Kölsch, H., & Kaiser, T. (2010). Reporting bias in medical research: A narrative review. *Trials, 11*, 37.

McQuay, H., & Moore, A. (1997). Using numerical results from systematic reviews in clinical practice. *Annals of Internal Medicine, 126*, 712–720. Retrieved from http://www.medicine.ox.ac.uk/bandolier/booth/painpag/NNTstuff/numeric.htm

Mendes, D., Alves, C., & Batel-Marques, F. (2017). Number needed to treat (NNT) in clinical literature: An appraisal. *BMC Medicine, 15*, 112. Retrieved from https://doi.org/10.1186/s12916-017-0875-8

Moher, D., Cook, D., Eastwood, S., Olkin, I., Rennie, D., & Stroup, F. (1999). Improving the quality of reports of meta-analyses of randomized controlled trails: The Quorom group statement. *Lancet, 354*(9193), 1896–1900.

Noblit, G., & Hare, R. (1988). *Meta-ethnography: Synthesizing qualitative studies*. Newbury Park, CA: Sage.

O'Brien, S., & Yi, Q. L. (2016). How do I interpret a confidence interval? *Transfusion, 56*(7), 1680–1683.

Oxford University Centre for Evidence-based Medicine. (2016, May). *The Oxford levels of evidence 2.1*. Retrieved from https://www.cebm.net/2016/05/ocebm-levels-of-evidence/

Pratt, J., Rhine, J., Smith, B., Stuart, C., & Greenwood, J. (1940). *Extra-sensory perception after sixty years*. New York: Henry Holt.

Rose, S., Bisson, J., Churchill, R., & Wessely, S. (2002). Psychological debriefing for preventing post traumatic stress disorder (PTSD). *Cochrane Database of Systematic Reviews, 2002*(2), CD000560. https://doi.org/10.1002/14651858.CD000560

Sandelowski, M., & Barroso, J. (2007). *Handbook for synthesizing qualitative research*. New York: Springer Publishing.

Scargle, J. (2000). Publication bias: The "file-drawer problem" in scientific inference. *Journal of Scientific Exploration, 14*(2), 94–106.

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

Sistrom, C., & Garvan, C. (2004). Proportions, odds, and risk. *Radiology, 230*(1), 12–19. https://doi.org/10.1148/radiol.2301031028

Smith, M., Glass, G., & Miller, T. (1980). *The benefits of psychotherapy*. Baltimore, MD: Johns Hopkins University Press.

Sridharan, L., & Greenland, P. (2009). Editorial policies and publication bias: The importance of negative studies (editorial commentary). *Archives of Internal Medicine, 169*, 1022–1023.

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 19*(3), 227–229.

Thombs, B., & Jewett, L. (2009). Letter: Analyzing effectiveness of long-term psychodynamic psychotherapy. *Journal of the American Medical Association, 301*(9), 930.