SECOND EDITION

COMPREHENSIVE SYSTEMATIC REVIEW *for* ADVANCED PRACTICE NURSING

Cheryl Holly
Susan Salmond
Maria Saimbert

EDITORS

SPRINGER PUBLISHING COMPANY

7 Critical Appraisal

Susan Salmond and Sallie Porter

OBJECTIVES

At the end of this chapter, the reader will be able to:

- *Articulate the steps in a critical appraisal*
- *Explain the importance of critical appraisal to the systematic review(SR) process*
- *Select the appropriate critical appraisal tool based on the study design*
- *Critically appraise a primary study or an SR*

CHAPTER HIGHLIGHTS

- *The quality of an SR is in part dependent upon the credibility of the primary studies that are included in the review.*
- *Critical appraisal is an assessment of the benefits and strengths of research along with its flaws and weaknesses.*
- *Critical appraisal is essential for the inclusion of the highest quality studies available for an SR.*
- *There are a number of different critical appraisal tools available for assessing the quality of research studies within the context of an SR based on the type of study design used.*
- *Two appraisers assess for risk of bias in each study to determine the study's internal validity and rigor. Use of two appraisers along with a third appraiser to negotiate opposing views improves accuracy and limits bias in the SR.*

Appraisal is an evaluation, judgment, or assessment. Critical appraisal is a process of systematically examining individual research studies to assess their reliability (or trustworthiness), worth (or value), and importance (or relevance) in a specific context. Each study selected for inclusion in an SR requires appraisal by at least two members of the review team. The appraisers work individually using pre-selected appraisal tools to determine study quality and appropriateness for use in SR findings and then compare their assessments (Holly & Porter, 2016).

An SR is a form of secondary research that gathers primary studies on a clinical or policy question of interest and analyzes the data from multiple studies to reach a conclusion. The quality of an SR is, to a large extent, dependent upon the credibility of the primary studies that are included in the review. To ensure this quality, critical appraisal is an integral part of the SR process (Hammick, Dornan, & Steinert, 2010).

Critical appraisal is completed independently by two well-qualified reviewers to assess for methodological rigor in order to determine whether the results of the primary research are sufficiently valid to be considered useful information (Evans, 2001). Reviewers who assess for methodological quality should be well versed in research design and analysis as reviewers decide based on the outcomes of the appraisal whether a study should be a part of the review or be excluded. A detailed log of the excluded studies is kept and in the final report, a table of excluded articles, providing the citation and the reason for exclusion, is included. In this way, the criterion of transparency in the review process is met.

Greenhalgh (2015) reminds us that the goal of critical appraisal is not finding methodologically flawless papers. With flaws in 99% of research studies, the aim is therefore to identify papers that are "good enough." The critical appraisal provides a balanced, scholarly assessment of the benefits and strengths of research against its flaws and weaknesses. The parameters used for evaluating validity vary according to the specific research design.

Three broad questions are addressed with quantitative critical appraisal—whether the results are valid, what are the results, and will the results help with one's own patient population or policy of interest. Validity refers to one's assessment of how close the study results are to reality. In studies where the aim is to determine cause and effect, the focus of validity appraisal is on internal validity. Thus, internal validity critical appraisal questions focus on assignment of participants to treatments, whether participants who entered the study are sufficiently accounted for at its conclusion, whether the groups were similar at the start of the trial, whether blinding was used, and whether groups were treated ethically. Appraisal focusing on quantitative results ensures there is adequate reporting of data collection methods, that analysis is appropriate, whether key findings are reported appropriately, and the significance and precision of the results. In quantitative research the focus is whether all important outcomes are reported and the benefits of the intervention versus the harms and the costs are considered.

Validity of qualitative research follows a different paradigm and the research is appraised for credibility of the researcher and credibility of methods. In qualitative research this focus is on confirmability and dependability. Applicability questions examine whether the results can be applied to one's own practice and population of interest. In qualitative research the aim is not to generalize but to transfer findings to situations that are contextually similar. Questions focus on the contextual similarity.

Tools for Critical Appraisal

There are many different tools available for assessing quality of research studies within the context of an SR. Tools vary based on the type of design used. These generally take the form of scales in which quality criteria are scored and combined in a summary score, or checklists in which specific questions are asked and the reviewers must determine which questions are critical to quality for inclusion. Checklists are the most commonly used and recommended. No single cut-off score is universally accepted, and validity has not been established for most of the appraisal approaches (Margaliot & Chung, 2007). If undertaking an SR for Cochrane reviews, Campbell reviews, or the Joanna Briggs Institute (JBI), there are specific appraisal tool requirements and the reader should refer to the appropriate website.

Papers need to be closely read and reread during the appraisal process. For novice appraisers or for novice review researchers, there is a learning curve to understanding and applying the appraisal criteria. For the novice, this can be facilitated by using a mentor for appraisal of a select number of articles until the mentor has deemed the researcher to have the necessary skills and competencies. Reviewer expertise in the content area of interest may assist in determining the usefulness of particular studies. As all appraisals are done by two reviewers with the availability of a third for areas of dispute, accuracy is enhanced.

Appraisal of Effectiveness Research

Randomized controlled trials (RCTs) are considered the gold standard for intervention research. RCTs usually measure short-term effects in select populations under strict (highly controlled) conditions—thus testing for efficacy. By using randomization or allocation by chance the intent is that the groups being compared are similar in terms of both measured and unmeasured baseline factors (Rochon et al., 2005), thus increasing the likelihood that differences in the dependent variable are attributable to the treatment variable. Critical appraisal of this type of design examines internal validity or the extent to which the study design, conduct, and analysis has minimized or avoided biases in its comparison of treatments. Bias is a form of systematic or predictable error in which the observed results may in fact be different from the true results. In the presence of significant bias, the results may not be considered valid or trustworthy. Therefore, in considering whether a study should be included in an SR of effectiveness, it is important to appraise for sources of bias. With RCTs, one would assess for bias in multiple domains: selection bias, performance bias, attrition bias, detection bias, and reporting bias (Higgins et al., 2011; JBI, 2014; Magarey, 2001). An explanation of these systematic biases and how to appraise for them is described further in Table 7.1.

There are dozens of scales and checklists that can be found for appraising effectiveness studies. Most of the tools measure beyond internal validity and

TABLE 7.1 APPRAISING FOR BIAS IN EFFECTIVENESS STUDIES

Type of Bias	Explanation of Bias	Critical Appraisal for Bias
Selection bias	■ Results from errors in the way that research participants were selected into the study from the target population or as a result of factors that influence whether research participants remained in a study. ■ The intervention group is therefore different from the control/comparison group in measured or unmeasured baseline characteristics and this difference may impact prognosis or outcomes. ■ Also used to mean that the participants are not representative of the population of all possible participants. ■ Greater chance for selection bias with nonrandom samples or when the individual assigning participants to intervention groups has the ability to select which group the individual will be assigned to.	Randomization and allocation concealment are key to minimizing selection bias. Evaluate whether: ■ Randomization was used ■ The allocation sequence was appropriate (such as using a random component in the sequence generation such as a random number table, coin toss, or throwing dice) ■ Allocation was adequately concealed
Performance bias	■ Systematic differences in care provided to the participants in the intervention and control/comparison group. ■ In the presence of differences in the care provided, cannot confidently conclude that the intervention under investigation caused the effect. ■ More likely to occur if the caregiver is aware of whether a patient is in a control or treatment group. ■ Blinding is an approach to prevent the subject and/or researcher clinician from knowing the allocated intervention.	Was there blinding of subject? Was there blinding of researcher/clinician?
Attrition bias	■ Differences between control and treatment groups in terms of patients dropping out of a study, or not being followed up as thoroughly as others in the groups. ■ Attrition of participants from a study can produce bias if the incidence rates in people who drop out differ from those in people who complete the study. ■ Although drop outs will occur, want to be assured that missing outcome data are balanced in numbers across groups with similar reasons for missing data across groups.	Was loss to follow up (i.e., dropout, nonresponse, withdrawal, protocol deviators) reported? Did researchers apply the concept of intention to treat?
Detection (assessor or ascertainment) bias	■ Occurs if outcomes are measured differently for patients depending on whether they are in the control or treatment group. ■ A detection bias generally occurs when the assessor (the one determining the outcome results) knows whether the subject is in the control or intervention group.	Was blinding of the assessor carried out?
Reporting	■ Occurs when outcomes are selectively reported in the findings. ■ Reporting bias may occur when the assessor does not fully report all positive or negative outcomes.	How were selective outcomes reported and what was found?
Other	■ Potential bias not uncovered in other domains.	Are there any other concerns identified that might indicate bias?

capture criteria of quality, study precision, and applicability. Questions such as Did the study ask a clearly focused question? Did the study have enough participants to minimize the play of chance? and How are the results presented and what is the main result? are examples of questions that capture the quality of the study rather than bias. Questions such as Are your patients so different from those studied that the results may not apply to them? target applicability. Questions such as Were participants appropriately allocated to intervention and control groups? Were participants, staff, and study personnel 'blind' to participants' study group? and Were all of the participants who entered the trial accounted for at its conclusion? are examples of questions focusing on potential bias.

Bias addresses the issue of believability of the findings. Quality focuses on whether the study was carried out to the highest possible standards and is larger in scope than bias. It is possible to have a study carried out with the highest possible standards yet still have an important risk of bias. For example, in a study where the participants are not blinded to their group allocation, the study may be carried out with high quality yet performance bias may still be an issue (Higgins & Green, 2008). The Cochrane group distinguishes bias from quality and has a separate tool to assess the risk of bias in RCTs (Higgins et al., 2011).

A challenge to the reviewer appraising studies is that review of the published articles often does not clearly and fully report the methodology. This may require contacting the corresponding author for clarification on the approaches used. The Consolidated Standards of Reporting Trials (CONSORT) Statement "is an evidence-based, minimum set of recommendations for reporting randomized trials" (CONSORT, n.d.). The CONSORT Statement has been endorsed by a number of editors of health care journals. A 25-item checklist of information to include when reporting an RCT as well as a flow diagram is available through CONSORT. Use of the CONSORT model in RCT reporting may assist those engaged in critical appraisal to better determine the risk of bias and quality of a study; however, the uptake of the CONSORT model has been incomplete and the overall quality of RCT reporting remains less than optimal (Chhapola, Tiwari, Brar, & Kanwal, 2015; Turner et al., 2012).

Appraisal of Observational Studies

Observational studies include a variety of types of designs (e.g., cohort, case-controlled) that compare outcomes in groups that did and did not receive an intervention, but allocation to intervention or non-intervention group is not through randomization. Observational studies are frequently the best choice in cases where ethics does not allow exposing a group to harmful agents, when measuring infrequent adverse outcomes, when evaluating interventions designed to prevent rare events, and when examining long-term outcomes (Song & Chung, 2010). As observational studies do not use randomization and generally have less restrictive inclusion and exclusion criteria as compared to RCTs, observational studies typically are at greater risk for bias, but are more reflective of

the population at large (external validity). One of the advantages of a cohort design is that it can determine whether efficacy observed in randomized trials translates into effectiveness in broader populations and in more realistic settings (Rochon et al., 2005).

When randomization is not used or when subjects can select their own treatments or their environments impose treatments upon them, there is a greater risk that the differences in outcomes are due to pretreatment differences (i.e., confounders) rather than to the effects of the treatment. CONSORT defines confounding as a situation in which the estimated intervention effect is biased because of some difference between the comparison groups apart from the planned intervention. This could include baseline characteristics, prognostic factors, or concomitant interventions (CONSORT, 2013). Selection of the comparison group should be done so that it is as close to matching the intervention group as possible and the decision making regarding comparison group selection should be clear.

To minimize this risk, researchers should use techniques such as matching to achieve comparability of the covariates of concern across groups. For example, in a case-control study where smoking is deemed a confounding factor, cases and controls can be matched by smoking status, so that for each case that smokes, a control who does not smoke is found (Boccia et al., 2007). Information on the distribution of potential confounders within the two comparison groups should be provided to demonstrate comparability. Restriction is another approach to control confounders. In restriction, one uses more stringent inclusion/exclusion criteria to limit the potential of a known confounder. However, there may be unknown confounders that would typically be evenly distributed in the presence of randomization but with matching there is less protection against potential confounders. Post hoc analysis can also minimize confounding effects. Use of stratification or multivariable modeling can provide estimates of the confounding effect.

As observational studies do not use randomization, appraisal focuses on the approach to recruitment of cases and controls in order to have matching or comparability of subjects, the accuracy of measurement of exposure and outcome to minimize bias, the identification of potential confounders, the careful definition of exposure for case selection, the precision of results, and the applicability to the local population. Although questions for reporting and applicability are generally similar to effectiveness studies, questions on validity will differ. Typical questions examining validity include whether there is support for the choice of the study method, whether the population studied is appropriate, whether confounding and bias are considered, and whether follow-up was long enough.

Appraisal of Qualitative Studies

Critical appraisal of qualitative studies differs significantly from critical appraisal of quantitative approaches with continuing debate as to whether critical appraisal

should be done at all for qualitative syntheses (Hannes, Lockwood, & Pearson, 2010). In this text we have adopted the viewpoint that critical appraisal is an integral part of the SR process with the goal to synthesize findings from high-quality studies. As qualitative research is grounded in a different philosophical perspective than quantitative research, the goal of critical appraisal for qualitative research is not to minimize bias, but rather determine the "reality" of results. Applying appraisal criteria from a quantitative paradigm to evaluate qualitative research is inappropriate and reviewers must use different criteria to judge the rigor of quantitative and qualitative designs.

DIFFERENCES IN THE QUANTITATIVE AND QUALITATIVE PARADIGMS

The quantitative paradigm is grounded in the concept of logical positivism with the belief that the world is made up of observable, measurable facts. The focus is on measurement and analysis of causal relationships between variables, prediction, and generalization of findings (Golafshani, 2003), and to this end the approach is reductionistic with an attempt to fragment and delimit phenomena into measurable or common categories. The focus is on defining concepts and operationalizing measures that represent the concepts. Thus, validity and reliability take on central importance in this paradigm (Devers, 1999).

Qualitative research uses a naturalistic approach and the goal is to understand phenomena within the real context, without any attempt to manipulate the phenomenon of interest. The qualitative research paradigm is grounded in constructivism or interpretivism, which views knowledge and meaning as dynamic, contextual, and socially constructed from the interaction between human beings and their world (Musthafa, 2014). The goal of qualitative research is illumination, generating understanding of phenomenon within specific contexts, and extrapolation to similar situations or contexts (Golafshani, 2003). In qualitative research, the researcher is the instrument and credibility of qualitative research depends on the ability and effort of the researcher.

CRITERIA FOR QUALITATIVE APPRAISAL

Although there are numerous appraisal tools for qualitative research, there is no clear cut definitive list of ideal criteria. The aim of critical appraisal is to ensure that there is enough breadth and depth of the data to suggest that the findings are trustworthy (Jones, 2004).

Cohen and Crabtree (2008) did a cross-publication content analysis of journal articles and texts that discussed vital criteria for rigorous qualitative research. They identified seven criteria for good qualitative research: (1) carrying out ethical research; (2) importance of the research; (3) clarity and coherence of the research report; (4) use of appropriate and rigorous methods; (5) importance of reflexivity or attending to researcher bias; (6) importance of establishing validity or credibility; and (7) importance of verification or reliability. They concluded that there was consensus on the first four criteria and divergent opinion

on the remaining three as qualitative researches have mixed views on whether it is appropriate for applying concepts such as bias, validity and reliability (even though the criteria have been adapted to be appropriate to the qualitative paradigm) to qualitative work.

There have been many attempts to render the concepts of validity, bias, and reliability to the qualitative paradigm. To this end the concepts of rigor, trustworthiness, plausibility, and credibility are the qualitative equivalent of validity, dependability, the corollary of reliability, and confirmability reflecting the process to decrease bias. The goal of qualitative work is not generalization; rather transferability, a term capturing the contextual similarity, is a more appropriate concept (Lincoln & Guba, 1985). Understanding these concepts is important to qualitative critical appraisal. Table 7.2 provides further explanation of these adapted concepts and strategies that can be used to secure them.

VALIDITY IN QUALITATIVE RESEARCH

Hannes et al. (2010) focused their efforts not on translation, but on clarifying what qualitative researchers do to establish validity and identification of the threats to validity. They define qualitative appraisal *criteria* as the standards to be upheld as ideals in qualitative research and *techniques* as the methods used to minimize threats to validity. Using Maxwell's framework, which stresses that validity is based on the kinds of understanding we have of the phenomena under study, the authors described and compared appraisal tools on five types of validity: descriptive, interpretive, theoretical, generalizable, and evaluative validity. They propose that descriptive, interpretive, and theoretical validity are central to qualitative research. It should be noted, however, that these categorizations are not discrete, but overlapping concepts.

Descriptive validity is "the degree to which descriptive information such as events, subjects, setting, time, and places are accurately reported" (Hannes et al., 2010, p. 5). It captures criteria that ask the reviewer to determine the impact of the investigator and whether the context was adequately reported. The manuscript should provide an understanding of the context in which the research was carried out. This helps in understanding the phenomenon as well as in decision making about transferability of the findings. Components of context include the physical setting as well as the investigator's role in the setting. As the researcher is the instrument, one must appreciate that the interaction of the researcher and setting influences the nature and types of data collected. Any events over time that could have changed the nature of the study or affected the results should be reported. To this end, appraisal criteria that address context could determine if there is a statement locating the researcher culturally or could ask the questions: What role does the researcher adopt within the setting? and Are the findings interpreted within the context of other studies and theories? Criteria focusing on the impact of the investigator could ask: Has the influence of the researcher on the research and vice versa been made clear? Has the relationship between researchers and participants been adequately considered?

TABLE 7.2 CRITICAL APPRAISAL FROM A QUALITATIVE PARADIGM

Concept	Definition/Explanation	Techniques to Facilitate Achieving the Concept
Credibility	■ The extent to which findings and conclusions are supported by data/study evidence. The findings/conclusion make sense and have a coherent logic. ■ Includes credibility of researcher, method, and findings.	Do the findings and conclusion make sense? Do the findings and conclusion seem logical?
Credibility of researcher	■ Researcher must be credible and any experiences that could influence interpretation of the phenomenon should be identified.	Are researchers' qualifications, experiences, perspectives, and assumptions identified? Are there any personal connections between the researcher and the topic or participants?
Credibility of method	■ Strategies can be built into data collection to enhance credibility of findings such as prolonged engagement for observation or interview methods. ■ Triangulation (using multiple methods or data sources) to study the phenomenon of interest.	Was there sufficient engagement with participants for the researcher to gather trustworthy data? Was more than one method used in data collection? Were the data sources used sufficient to gather a full perspective of the phenomenon of interest?
Credibility of findings	■ Use of a second researcher to analyze and confirm data and to ask questions about methods, meanings, and interpretation of the data. ■ Having participants from the study review findings and give their views as to the credibility of the interpretations and findings enhances credibility. ■ Triangulation uses multiple data sources, investigators, methods, or theory to provide corroborating evidence to substantiate the findings. ■ Search for disconfirming evidence ("deviant" or "negative" cases): Actively search for cases that do not fit the pattern and refine the theory and working hypotheses in the light of this evidence.	Was a second researcher used for peer review? Was peer debriefing used? Was member checking used? Was triangulation used? Were approaches to seek disconfirming evidence identified?
Transferability (also referred to as applicability and fittingness)	■ Transferability addresses the extent to which the findings can be applied to other contexts. ■ Thick description and purposive sampling are strategies to enhance transferability. ■ Providing a detailed description of the context facilitates decisions about transferability.	Was the sample and context clearly defined?

(*continued*)

TABLE 7.2 CRITICAL APPRAISAL FROM A QUALITATIVE PARADIGM (*continued*)

Concept	Definition/Explanation	Techniques to Facilitate Achieving the Concept
Dependability (reliability)	■ Achieved through the use of an audit trail or data archiving; allows an independent examiner to track the decisions made and steps taken in the study. ■ Use of a skeptical peer review where another individual skilled in the research approach asks questions about the methods, meanings, and interpretation of the data.	Did the study describe methods to identify transparency in decision making during the analysis process? Was the analysis confirmed by a second researcher? Was a tape recorder or other mechanical device used to record the interviews?
Confirmability (objectivity, bias reduction)	■ Use of triangulation. ■ Use of skeptical peer review or audits. ■ Search for disconfirming evidence or negative cases. ■ Reflective journal keeping by the researcher chronicling how his or her personal characteristics, feelings, and biases may be influencing the work and what strategies have been used to manage them.	Was triangulation used? Did the study describe methods to identify transparency in decision making during the analysis process? Was the analysis confirmed by a second researcher? Did the researcher document a reflexive approach?

Sources: Beck (1993); Devers (1999); Patton (2002).

Are the researcher's own position, assumptions, and possible biases outlined? and Is there evidence of reflexivity—that the researcher has reflected on his or her potential personal influence in the collection and analysis of data?

Interpretive validity is "the degree to which participant's viewpoints, thoughts, intentions, and experiences are accurately understood and reported by the qualitative researcher" (Hannes et al., 2010, p. 5). Using the words of the informants as building blocks, researchers construct interpretive accounts of the phenomenon. The primary question is the "believability" of this account; does it "ring true"? Appraisal criteria may ask Are participants, and their voices heard? or Is there adequate evidence provided to support the analysis? Techniques to enhance interpretive validity include member checking, participant feedback, close collaboration with participants, peer debriefing, methods and analytic triangulation, and self-reflection by the researcher on potential biases, preconceptions, assumptions, and reference frameworks (Hannes et al., 2010).

Theoretical validity is "the degree to which a theory or theoretical explanation informing or developed from a research study fits the data and is, therefore, credible and defensible" (Hannes et al., 2010, p. 5). Criteria capturing theoretical validity include items on the theoretical framework (e.g., Is there congruency between the stated philosophical perspective and the research methodology? What theoretical framework guides or informs the study?) and on evaluation/outcome criteria (e.g., Do conclusions drawn in the research report appear to flow from the analysis, or interpretation, of the data? or Is the

conclusion justified given the conduct of the study?). Techniques to enhance theoretical validity include prolonged engagement such that there is "sufficient time to study the subjects and setting and to create a set of patterns and relationships that are stable and contribute to an understanding of why these occur" (Hannes et al., 2010, p. 6). Theory triangulation and searching for disconfirming or negative cases are also approaches to ensuring theoretical validity.

Generalizability (external validity) is "the degree to which findings can be extended to other persons, times, or settings than those directly studied" (Hannes et al., 2010, p. 5). In appraisal tools the criterion addressing this concept falls under the criterion value and implications of research. Items identified assessing this criterion include: How valuable is the research? To what setting and population are the study findings generalizable? and What are the implications for policy and practice?

Evaluative validity is "the degree to which an evaluative framework or critique is applied to the object of the study" (Hannes et al., 2010, p. 5). It establishes the "degree to which a certain phenomenon under study is legitimate, justified, or raises questions, and involves the application of an evaluative framework to the phenomenon under study" (Hannes et al., 2010, p. 3). Critical appraisal criteria that have an item on "outcome/evaluation" may capture a component of evaluative validity; however, most qualitative researchers do not generally attempt to evaluate the phenomenon under study.

EVALUATION OF THE QUALITY OF STUDY DESIGN

As with all research, the research design should be appropriate for the question of interest. Sampling approaches, data collection methods, data types and sources, data analysis methods, and reporting of findings are criteria that can be used to facilitate an overall judgment of the quality of the research and whether the researchers conducted the study consistent with standards of qualitative research (Hannes et al., 2010). Qualitative analysis, unlike quantitative analysis, does not follow a strict formula and rules-oriented approach to analysis; rather, it uses a more creative process that depends on the insights and conceptual capabilities of the analyst (Patton, 1999). It requires the researcher to recognize patterns, which in part is an intuitive process. However, complementing the intuitive process is technical rigor, which is a systematic process. The qualitative researcher needs to report sufficient details related to data collection and the processes of analysis used in order for readers to judge the quality of the resulting product. In qualitative research, designs evolve during data collection and analysis. The study report should clearly indicate how and why the research design changed.

STRATEGIES FOR ENHANCING RIGOR

There are several strategies that can be employed by qualitative researchers to enhance study rigor and methods for gathering high-quality data that are

carefully analyzed and consequently provide increased confidence in the findings. Some of these strategies have been presented in Table 7.2 but will be expanded upon in the text that follows.

Triangulation is an approach that corroborates that the research findings accurately reflect people's perceptions. There are different types of triangulation: researcher, method, source, and theoretical. Using multiple observers/researchers recording and describing the participants' behavior and context allows for cross checking of observations and enhances descriptive validity. Using multiple methods broadens the understanding of the phenomenon. Its usefulness arises from the logic that no single method ever adequately addresses the problem of rival explanations. As each method reveals different aspects of empirical reality, multiple methods of data collection and analysis provide cross-data validity checks enhancing confidence in the trustworthiness of the findings/explanations and, therefore, enhance interpretive validity. Examining the consistency of different data sources within the same method is triangulation of sources. Theoretical triangulation uses multiple perspectives or theories to interpret and explain the data. Patton reminds us, however, that the goal is not to find the same result, but to test for consistency in the results. Different kinds of data may yield somewhat different results as different modes of inquiry are sensitive to different real-world nuances and better understanding of these inconsistencies is illuminative (Patton, 1999).

Analysis or documentation of evaluator effects is considered important to establishing credibility or validity of qualitative research. There are four ways in which the presence of the researcher can influence the findings of a study: (1) the participants may react differently due to the presence of the researcher; (2) the researcher may change during the course of the data collection or analysis; (3) the researcher brings forward predispositions, selective perceptions, and/or biases; and (4) the researcher's level of training or preparation may influence the ability to collect and analyze the data (Patton, 1999, 2002). Not only must these areas of concern be addressed in the study but there are collection and analysis techniques that minimize this threat. Ensuring that there is adequate or prolonged engagement allows for a period of time for researcher and participants to get used to each other. Keeping daily field notes and monitoring observer effects by either direct questioning or observation demonstrates the researcher's awareness of this threat. Similarly, daily field notes with attention to shifts in one's own attitudes and behaviors monitors researcher change. Addressing researcher predispositions or biases involves declaring any personal connections between the researcher and the topic or participants, stating up front prior interpretations or experiences with the phenomenon, as well as maintaining a posture of empathic neutrality in which the researcher is caring toward the people under study but neutral or impartial about the findings. Lastly, the issue of researcher competence is closely tied to perceived credibility of the findings. The study should document what verification or validation procedures were used to establish quality of the analysis. Use of an

audit trail or data archiving allows for an independent examiner to track the decisions made and steps taken in the analysis of the data. Having a "skeptical peer review" where another researcher skilled in the research approach asks questions about the methods, meanings, and interpretation of the data, or having an independent researcher analyze samples of the data to compare interpretation are techniques to enhance the trustworthiness of the findings. The analysis process and findings in terms of patterns, linkages, and plausible explanations should be clearly described and the report should provide information on approaches taken to test these findings by searching for disconfirming evidence or negative cases.

Transferability is the degree to which the findings from qualitative research can be transferred or generalized to other contexts or settings. Transferability can be enhanced by thoroughly describing the research context (i.e., study participants, demographics, contextual background information) and providing thick description on the sending and receiving context. In this way readers can make informed decisions about "to whom the results might be generalized or to which groups the findings can be transferred" (Hannes et al., 2010, p. 6).

Critical Appraisal of SRs

The purpose of an SR is to provide reliable evidence summaries. The quality of the review and what tends to set SRs apart from narrative reviews is in part based on the degree to which systematic methods were used to reduce the risk of error and bias. An SR becomes systematic through development and adherence to an explicit and auditable protocol for review (O'Mathúna, Fineout-Overholt, & Kent, 2008; Sandelowski, 2008). The protocol, as described in Chapter 3, sets forth the background for the review, the review question (generally using the PICO or PICo formats [see next page]), the inclusion and exclusion criteria that will guide study selection, the search strategy, approach to appraisal of the studies, and the approach to analyze and synthesize the findings. Although a published manuscript may not provide the total protocol/proposal, it should at least be referred to in the published manuscript giving evidence of an a priori plan for the process.

Even when using an established protocol, the reality is that not all SRs are of equal quality and it is critical for the user of research to critically appraise the SR for its degree of rigor. As with single studies, there are many credible approaches to critical appraisal of SRs. Used here is an adaptation of the JBI guide for systematic review appraisal (2000). The questions, described in Table 7.3, follow the general review process, ending with reporting of the findings and conclusions and recommendations.

The *research question* should be clearly and explicitly stated. If not clearly stated, the utility of the remainder of the study is questionable. An SR tends to have narrow or focused questions but many SRs have been broader using

TABLE 7.3 CRITICAL APPRAISAL OF A SYSTEMATIC REVIEW

Review Question	Is the Review Question Clearly and Explicitly Stated?
Search strategy	Were comprehensive search methods used to locate studies? Was a thorough search done of appropriate databases and were other potentially important sources explored?
Inclusion criteria	How were studies selected?
Critical appraisal	Was the validity of studies assessed appropriately?
Similarity of studies	Were the populations of the different studies similar? Was the same intervention/phenomenon of interest evaluated by the individual studies? Were the same outcomes used to determine the effectiveness of the intervention being evaluated? Were reasons for differences between studies explored?
Reporting of findings	Are review methods clearly documented? Is the review question clearly and explicitly stated? Was the search strategy reported? Was the inclusion criteria reported? Was the criteria for appraising studies reported? Were the methods used to combine studies reported?
Conclusions and recommendations	Is a summary of findings provided? Are specific directives for new research proposed? Were the recommendations supported by the reported data?

Reprinted with permission from the Joanna Briggs Institute (2000).

several subquestions. Using the PICO format in quantitative research, the question encompasses the components of (a) population of interest and condition; (b) intervention; (c) comparison or control intervention; and, (d) outcome of interest. Setting and time may also be part of the question. For qualitative reviews, the mnemonic PICo captures the components of (a) population; (b) phenomena of interest; and (c) context.

A clear, transparent and comprehensive *search strategy* is a critical component of an SR (Manchikanti, 2008; O'Mathúna et al., 2008). The search strategy should be transparent so that it is reproducible, not only by other researchers but by the primary author so that the review can be updated. The aim of a comprehensiveness search is to locate as much of the completed research on the topic as possible (JBI, 2000).This involves not only searching multiple databases and the grey literature, but also using additional search strategies such as footnote chasing and key author and organization identification to identify missed papers—both published and unpublished. Appraisal of the search strategy captures whether the appropriate bibliographic databases were used, if there were follow-up from reference lists, if personal contact with experts were made, if unpublished studies were identified, and if non-English language studies were searched.

Inclusion and exclusion criteria operationalizes the review question and reduces selection bias by specifying a priori the boundaries rather than including/excluding studies based on their results. These criteria define the scope of the review and should specify which type of study designs will be included in the review, as well as specifying the parameters for the population, intervention, and outcomes. Additionally temporal and linguistic constraints should be identified (Manchikanti, 2008). By establishing these criteria a priori it limits the risk of selection bias.

As clearly highlighted earlier in this chapter, in the SR process *critical appraisal* is done of each potential study's research design to ensure the validity prior to the final determination to include or exclude the study. As the goal is to pool data, the purpose of critical appraisal is to exclude lesser quality studies, thereby minimizing the risk of error and bias in the SR findings. Thus, when appraising an SR, it is necessary to determine whether the validity of studies was assessed appropriately. Appraisal should have been completed by two reviewers. Appraisal criteria vary by research design, but the review criteria used should be clearly documented. The review should provide a table of included and excluded studies. For included studies an aggregated table of characteristics should be provided.

Reporting of findings or data synthesis provides the summary of results from the different studies in order to obtain an overall evaluation of the effectiveness of an intervention or a more in-depth representation of a phenomenon of interest. Quantitative results are reported in terms of treatment effects and precision. An overall evaluation of the effectiveness of an intervention or treatment can be determined with a meta-analysis. A meta-analysis takes similar measures from comparable studies and when possible the measures of the effect are combined. Methods for combining the studies should be reported. Studies are not pooled when there are differences in the population, intervention, or how outcomes were measured, or when findings differ significantly. Thus, the reports should specify whether the populations of the different studies were similar; whether the same interventions were evaluated; whether the same outcomes were used to determine effectiveness; and whether the reasons for differences between studies were explored. The characteristics and results of each study should be clearly displayed in an included studies table. When pooling is not possible, a narrative table can be used presenting focused results. For qualitative SRs the process of data extraction should be clearly documented with a clear depiction of how findings were interpreted into a new coding structure and synthesis.

Conclusions, recommendations, and implications for research and clinical practice flow from the findings of the review. For effectiveness studies, the main result and the size and confidence of that result guide recommendations and implications. For all designs, the authors should provide a summary of the findings, suggest new directions for research, and make recommendations for practice that are supported by the reported data.

Practice Activities

1. Select a journal article about a research study that was analyzed using statistical techniques. Appraise the article using CATmaker. CATmaker is a software program, downloadable from http://www.cebm.net/catmaker-ebm-calculators, designed to help you appraise a paper and to manipulate the outcome statistics into a useful and standard format. It will help with most types of journal articles.
2. Go to the International Center for Allied Health Evidence (ICAHE) website at http://www.unisa.edu.au/cahe

   Review the variety of critical appraisal tools that are currently in use. Decide which you would use if you were doing an SR on: (a) interventions to decrease pressure ulcers in nursing homes; (b) strategies to retain nurses; (c) perceptions of patients regarding the effectiveness of over the counter pain medication; (d) experiences of caregivers of those with dementia; and (e) characteristics and risk factors of acute delirium in hospitalized children.

Suggested Reading

Boccia, S., La Torre, G., Persiani, R., D'Ugo, D., van Duijn, C.M., & Ricciardi, G. (2007). A critical appraisal of epidemiological studies comes from basic knowledge: A reader's guide to assess potential for biases. *World Journal of Emergency Surgery, 2*, 7.

Clegg, A., Hewitson, P., & Milne, R. (2000). Explicit and reproducible: How to assess the quality of the evidence in a systematic review. *Student British Medical Journal*, *8*, 24.

Hannes, K., Lockwood, C., & Pearson, A. (2010). A comparative analysis of three online appraisal instruments' ability to assess validity in qualitative research. *Qualitative Health Research*, *20*(12), 1736–1743.

References

Beck, C. T. (1993). Qualitative research: The evaluation of its credibility, fittingness and auditability. *Western Journal of Nursing Research*, *15*(2), 263–266.

Boccia, S., La Torre, G., Persiani, R., D'Ugo, D., van Duijn, C. M., & Ricciardi, G. (2007). A critical appraisal of epidemiological studies comes from basic knowledge: A reader's guide to assess potential for biases. *World Journal of Emergency Surgery*, *2*, 7.

Chhapola, V., Tiwari, S., Brar, R., & Kanwal, S. K. (2015). Impact of "CONSORT extension for abstract" publication on the reporting quality of abstracts of randomized controlled trial in high impact journals of pediatrics—an interrupted time-series analysis. *Journal of Clinical Epidemiology*, *69*, 324–333. doi:10.1016/j.jclinepi.2015.10.013

Cohen, D. J., & Crabtree, B. F. (2008). Evaluative criteria for qualitative research in health care: Controversies and recommendations. *Annals of Family Medicine*, *6*(4), 331–339.

Consolidated Standards of Reporting Trials. (2013). Glossary items: Confounder. Retrieved from http://www.consort-statement.org/Search?q=confounder

Consolidated Standards of Reporting Trials. (n.d.) The consort statement. Retrieved from http://www.consort-statement.org/

Devers, K. J. (1999). How will we know "good" qualitative research when we see it? Beginning the dialogue in health services research. *Health Services Research*, *34*(5), 1153–1188.

Evans, D. (2001). Systematic reviews of nursing research. *Intensive and Critical Care Nursing*, *17*(1), 51–57.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *Qualitative Report*, *8*(4), 597–607.

Greenhalgh, T. (2015). *How to read a paper* (5th ed.). Oxford, UK: Blackwell.

Hammick, M., Dornan, T., & Steinert, Y. (2010). Conducting a best evidence systematic review. *Medical Teacher*, *32*, 3–15.

Hannes, K., Lockwood, C., & Pearson, C. (2010). A comparative analysis of three online appraisal instruments' ability to assess validity in qualitative research. *Qualitative Health Research*, *20*(12), 1736–1743. doi:10.1177/1049732310378656

Higgins, J. P. T., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions*. West Sussex, England: Wiley-Blackwell.

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., . . . Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal, 343*, d5928. Retrieved from http://doi.org/10.1136/bmj.d5928

Holly, C., & Porter, S. (2016). Appraising and extracting data, Chapter 4. In Y. Jadotte, C. Holly, S. Salmond (Eds.), *Workbook for systematic review* (pp. 20–26). Philadelphia, PA: Lippincott.

Joanna Briggs Institute. (2000). Appraising systematic reviews. *Changing Practice: Evidence Based Practice Information Sheets for Health Professionals. Supplement 1*, 1–6.

Joanna Briggs Institute. (2014). *Joanna Briggs Institute reviewer's manual: 2014 Ed*. Adelaide, Australia: Author. Retrieved from http://joannabriggs.org/assets/docs/sumari/ReviewersManual-2014.pdf

Jones, M. L. (2004). Application of systematic review methods to qualitative research: Practical issues. *Journal of Advanced Nursing*, *48*(3), 271–278.

Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Magarey, J. M. (2001). Elements of a systematic review. *International Journal of Nursing Practice*, *7*(6), 376–382.

Manchikanti, L. (2008). Evidence-based medicine, systematic reviews, and guidelines in interventional pain management, part I: Introduction and general considerations. *Pain Physician*, *11*(2), 161–186.

Margaliot, Z., & Chung, K. C. (2007). Systematic reviews: A primer for plastic surgery research. *Plastic and Reconstructive Surgery*, *120*(7), 1834–1841.

Musthafa, M. N. (2014). Pluralism as multi-method scholarship0 in qualitative social science research. *IOSR Journal of Research & Method in Education*, *4*(5), 5–14.

O'Mathúna, D. P., Fineout-Overholt, E., & Kent, B. (2008). How systematic reviews can foster evidence-based clinical decisions: Part II. *Worldviews on Evidence-Based Nursing*, *5*(2), 102–107.

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research*, *34*(5), 1189–1208.

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.

Rochon, P. A., Gurwitz, J. H., Sykora, K., Mamdani, M., Streiner, D. I., Garfinkel, S., . . . Anderson, G. M. (2005). Reader's guide to appraisal of cohort studies: Role and design. *British Medical Journal*, *330*(7496), 895–897.

Sandelowski, M. (2008). Reading, writing and systematic review. *Journal of Advanced Nursing*, *64*(1), 104–110.

Song, J. W., & Chung, K. C. (2010). Observational studies: Cohort and case-control studies. *Plastic Reconstructive Surgery*, *126*(6), 2234–2242.

Turner, L., Shamseer, L., Altman, D. G., Weeks, L., Peters, J., Kober, T., . . . Moher, D. (2012). Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomized controlled trials (RCTs) published in medical journals. *Cochrane Database Systematic Reviews*, *11*, MR000030.