Evaluation of qualitative research studies

Clinical scenario
You work on a palliative care unit where you have many opportunities to discuss end of life decisions with patients and family members. In a recent team meeting of your unit’s providers, the topic of “appropriate” treatment choices for patients at end of life comes up. Some providers believe that they should counsel patients and family members to “help them make better end of life decisions so that they will have a good death.” There is, however, no consensus about how this should be done.

Finding the evidence
You volunteer to see if any studies have been done on decision making at the end of life. You remember that your institution has an online subscription to Evidence-Based Nursing. You sign in and go to the search screen. In the field “word(s) anywhere in article” you type in “end of life” (in quotations because you are looking for articles that include all 3 words together) and “decisions”. 4 matches are found. The first is an abstract entitled “Providers tried to help patients and families make end of life decisions so that they will have a good death.” There is, however, no consensus about how this should be done.

Is the research question clear and adequately substantiated?
Before proceeding with a full fledged review of the study, readers should look for the precise question the study sought to answer and consider its relevance to their own clinical questions. The study report should clearly document what is already known about the phenomenon of interest.

Is the design appropriate for the research question?
More than 40 unique approaches to qualitative research methods have been identified.17 Common approaches in published healthcare research include ethnography, grounded theory, and phenomenology. Other approaches include case studies, narrative research, and historical research. Traditional ethnography seeks to learn about culture from the people who actually live in that culture.18 A grounded theory approach is used to discover the social-psychological processes inherent in a phenomenon,19 whereas a phenomenological approach is used to gain a deeper understanding of the nature or meaning of the everyday “lived” experiences of people.20 References to articles that describe these different qualitative research approaches in greater detail are listed in the online version of this Users’ guide.

Qualitative approaches arise from specific disciplines and are influenced by theoretical perspectives within those disciplines. A critical analysis of a qualitative study considers the “fit” of the research question with the qualitative method used in the study.21 Although the specific criteria for proper application of each methodological approach vary somewhat, there are sufficient similarities among the approaches to discuss them in general.

Was the method of sampling appropriate for the research question and design?
The emergent nature of qualitative research that results from the interaction between data collection and data analysis requires that investigators not prespecify a sample for data collection in strict terms. Instead, important data sources be overlooked. In quantitative studies, the ideal sampling standard is random sampling. Most qualitative studies use purposeful (or purpose) sampling, a conscious selection of a small number of data sources that meet particular criteria. The logic and power of purposeful sampling lie in selecting information rich cases (participants or settings) for indepth study to illuminate the questions of interest.23 This type of sampling usually aims to cover a range of potentially relevant social phenomena and perspectives from an

Questions to help critically appraise qualitative research

Are the findings valid?
Qualitative researchers do not speak about validity in the same terms as quantitative researchers. In keeping with the world views and paradigms from which qualitative research arises, validity, or whether the research reflects best standards of qualitative science, is described in terms of rigour, credibility, trustworthiness, and believability. Numerous articles and books focus on validity issues for qualitative research.24–26 Similarly, there are several qualitative research designs, and each has slightly different conventions for their appropriate conduct. This Users’ guide provides an overview of the critical appraisal of qualitative research but, as with various quantitative research designs, there are variations in how rigour and validity are addressed in specific designs.

Are the findings of this study valid?
1. Is the research question clear and adequately substantiated?
2. Is the design appropriate for the research question?
3. Was the method of sampling appropriate for the research question and design?
4. Were data collected and managed systematically?
5. Were the data analysed appropriately?

What are the findings?
1. Is the description of findings thorough?

How can I apply the findings to patient care?
1. What meaning and relevance does the study have for my practice?
2. Does the study help me understand the context of my practice?
3. Does the study enhance my knowledge about my practice?
appropriate array of data sources. Selection criteria often evolve over the course of analysis, and investigators return repeatedly to the data to explore new cases or new perspectives.

Readers of qualitative studies should look for sound reasoning in the description and justification of the strategies for selecting data sources. Patton offers a succinct, clear, and comprehensive discussion of the various sampling strategies used in qualitative research.10 Convenience sampling is one of the most commonly used, yet one of the least appropriate, sampling strategies. In convenience sampling, participants are primarily selected on the basis of ease of access to the researcher and, secondarily, for their knowledge of the subject matter. Purposive non-probability sampling strategies include (1) judgmental sampling, where theory or knowledge points the researcher to select specific cases; (a) maximum variation sampling, to document range or diversity; (b) extreme or deviant case sampling, where it is necessary to select cases that are unusual or special in some way; (c) typical or representative case sampling, to describe and illustrate what is typical and common in terms of the phenomenon of interest; (d) critical cases, to make a point dramatically; and, (e) criterion sampling, where all cases that meet some predetermined criteria are studied (this sampling strategy is commonly used in quality improvement); (2) opportunistic sampling, where availability of participants guides on-the-spot sampling decisions; (3) snowball, network, or chain sampling, where people nominate others for participation; and (4) theory based operational construct sampling, where incidents, time periods, people, or other data sources are sampled on the basis of their potential manifestation or representation of important theoretical constructs. Participant observation studies typically use opportunistic sampling strategies, whereas grounded theory studies use theory based operational construct sampling.

Sample size is a critical question for all research studies. A study that uses a sample that is too small may have unique and particular findings such that its qualitative transferability or quantitative generalisability becomes questionable. In qualitative research, however, even studies with small samples may help to identify theoretically provocative ideas that merit further exploration. Studies with samples that are too large are equally problematic. Whereas quantitative research has specific guidelines that frame researchers’ decisions about adequate sample size, there are only general principles, reflective of judgment and negotiation, for qualitative researchers. Examination of several areas will help readers to identify the adequacy of sample size in qualitative studies. Firstly, references about the specific method used may offer some guidance. For example, sample sizes in phenomenological studies are typically smaller than those in grounded theory and ethnographic studies. Secondly, the trade-off between breadth and depth in the research affects sample size. Studies with smaller samples can more fully explore a broader range of participants’ experiences, whereas studies with larger samples typically focus on a more narrow range of experiences. Thirdly, readers can review published studies that used similar methods and focused on similar phenomena for guidance about sample size adequacy. Qualitative researchers judge the adequacy of a sample for a given study by how comprehensively and completely the research questions were answered. Readers of qualitative studies are encouraged to review the researcher’s documentation of sample size and selection throughout the course of the study.

Were data collected and managed systematically?
Qualitative researchers commonly use one or more of 3 basic strategies for collecting data. One strategy is to witness events and record them as they occur (field observation). Another strategy is to question participants directly about their experience (interviews). Finally, researchers may review written material (document analysis). Readers should consider which data collection strategies researchers used and whether these strategies would be expected to offer the most complete and accurate understanding of the phenomenon.

Regardless of the strategy, the approach to data collection must be comprehensive to avoid focusing on particular, potentially misleading aspects of the data. Several aspects of a qualitative report indicate how extensively the investigators collected data: the number of observations, interviews, or documents; the duration of the observations; the duration of the study period; the diversity of units of analysis and data collection techniques; the number of investigators involved in data collection and analysis; and the degree of investigators’ involvement in data collection and analysis notes.11–13 Taping and transcribing interviews (or other dialogue) is often desirable, but not necessary for all qualitative studies.

Were the data analysed appropriately?
Qualitative researchers often begin with a general exploratory question and preliminary concepts. They then collect relevant data, observe patterns in the data, organise these into a conceptual framework, and resume data collection to both explore and challenge their developing conceptualisations. This cycle may be repeated several times. The iterations among data collection and data interpretation continue until the analysis is well developed and further observations yield redundant, minimal, or no new information to further challenge or elaborate the conceptual framework or in-depth descriptions of the phenomenon (a point often referred to as saturation or informational redundancy). This “analysis stopping” criterion is so basic to qualitative analysis that authors seldom declare that they have reached this point; they assume readers will understand.

In the course of analysis, key findings may also be corroborated using several information sources, a process called data triangulation. Triangulation is a metaphor and does not mean literally that 3 or more sources are required. The appropriate number of sources depends on the importance of the findings, their implications for theory, and the investigators’ confidence in their validity. Because no 2 qualitative data sources will generate exactly the same interpretation, much of the art of qualitative interpretation involves exploring why and how different information sources yield slightly different results.14–16 Readers may encounter several useful triangulation techniques for validating qualitative data and their interpretation in the analysis.17–22 Investigator triangulation requires that > 1 investigator collect and analyse the data, such that the findings emerge through consensus between or among investigators. This is typically accomplished by an investigative team. Inclusion of team members from different disciplines helps to prevent personal or disciplinary biases of a single researcher from excessively influencing the findings. Theory triangulation is a process whereby emergent findings are examined in relation to existing social science theories.23–25 It is conventional for authors to report how their qualitative findings relate to prevailing social theory, although some qualitative researchers suggest that such theories should not be used to guide the research design or analysis.

Some researchers seek clarification and further explanation of their developing analytic framework from study participants, a step known as member checking. Most commonly, researchers specify that member checking was done to inquire whether participants’ viewpoints were faithfully interpreted, to determine whether there are gross errors of fact, and to ascertain
whether the account makes sense to participants with different perspectives.

Some qualitative research reports describe the use of qualitative analysis software packages. Readers should not equate the use of computers with analytic rigour. Such software is merely a data management tool for efficiently storing, organising, and retrieving qualitative data. These programs do not perform analyses. The investigators do the analysis as they create the keywords, categories, and logical relations used to organise and interpret the electronic data. The soundness of qualitative study findings depend on investigator judgments, which cannot, as yet, be programmed into software packages.

We indicated earlier that qualitative data collection must be comprehensive (ie, adequate in its breadth and depth) to yield a meaningful description. The closely related criterion for judging whether data were analysed appropriately is whether this comprehensiveness was determined in part by the research findings, with the aims of challenging, elaborating, and corroborating the findings. This is most apparent when researchers state that they alternated between data collection and analysis, collected data with the purpose of elucidating the “analysis in progress”, collected data until analytic saturation or redundancy was reached, or triangulated findings using any of the methods mentioned.

What are the findings?
Is the description of findings thorough?
Qualitative researchers are challenged to make sense of massive amounts of data and transform their understandings to a written form. The written report is often a barrier to qualitative research. Such reports are because of its lack of clarity and relevance, except to a limited audience. Sandelowski describes the challenges facing authors, as they make decisions in balancing description (the facts of the cases observed) with analysis (the breakdown and recombining of data) and interpretation (the new meanings created from this process).

Good research often involves “messiness”, raising as many questions as it purports to answer. Holliday describes the appropriate role of “cautious detachment” in qualitative research. The “truths” of qualitative research are relative to the research setting. Therefore, it is important that authors not overstep the interpretive boundaries of their study by making it seem as if all their questions were answered with certainty and without raising additional questions. A comparison of the findings and discussion sections of a study report is helpful for judging whether authors are truthful to the data and the local context of a given qualitative study.

How can I apply the findings to patient care?
What meaning and relevance does the study have for my practice?
Thorne suggests that critiquing qualitative research in health sciences disciplines demands not only a focus on traditional appraisal criteria, but also an examination of the more complex question of what meaning can be made of the findings. The moral question of how research findings may be used in ways not intended and not benefiting health science disciplines and patients is an important one, given that “health science disciplines exist because of a social mandate that entails a moral obligation toward benefiting individuals and the collective”.

Thorne describes 5 criteria for appraising the disciplinary relevance and usefulness of a study: (1) Are there convincing claims about why this knowledge is needed (moral defensibility)? (2) Is the knowledge appropriate to the development of the discipline (disciplinary relevance)? (3) Does the study produce usable knowledge (pragmatic obligation)? (4) Is the study situated in a historical context and within a disciplinary perspective (contextual awareness)? and (5) Is there evidence of ambiguity and creation of meaning (probable truth)?

Does the study help me understand the context of my practice?
The context in which a study is done influences the results of all research, but it is particularly important in qualitative research. Readers of qualitative research must determine the potential applicability of the findings to their own contexts. Inadequate reporting of the social and historical context of a study makes it difficult for readers to determine if a study’s results can be “transferred” with any legitimacy to their situation.

Does the study enhance my knowledge about my practice?
One criterion for the generalisability of a qualitative study is whether it provides a useful map for readers to understand and navigate in similar social settings themselves. Readers need to consider the similarity of the patients and setting of a given study to their own.

Resolution of the clinical scenario
You begin your critical appraisal of the study by Norton and Bowers by applying the criteria described above.

Is the research question clear and adequately substantiated? and Is the design appropriate for the research question?
Norton and Bowers explored how providers described their work in changing patients’ and families’ treatment decisions at end of life from what providers deemed curative to palliative (unrealistic to more realistic). The stated purpose of the study was “to develop a grounded theory of how decisions were negotiated among providers and family members near the end of a patient’s life”. They described how, during the development of the grounded theory, they identified “several strategies providers used to assist patients and families to shift from curative to palliative treatment choices and goals”. The study report focused on those strategies. The authors clearly stated that this report focused on one portion of a larger grounded theory that was derived from a larger main study.

Norton and Bowers discuss background literature on patient self determination, advance directives, level of treatment received, beliefs about prognosis, changes over time of patient treatment decisions, and how patients, families, and providers achieve agreement on treatment decisions.

The use of a grounded theory method was appropriate for this study, given that the authors were interested in the meanings that providers attributed to end of life treatment choices of patients and families and how providers attempted to shift patients’ and families’ understandings of the “big picture” to influence their treatment decisions.

Was the method of sampling appropriate for the research question and design?
Norton and Bowers interviewed 15 healthcare providers. Given that theoretical sampling is a key sampling strategy in grounded theory research, the authors discussed how they altered the design of the interviews to identify whether providers assessed patients’ and families’ understanding, whether they used strategies to help patients and families come to a more realistic understanding of their situations, and how providers understood their actions and what they were trying to accomplish. As
the research progressed, Norton and Bowers described theoretical sampling of types of providers (nurses and physicians), work settings (home health, family practice, oncology, and intensive care), and work experience (experienced or novice, in terms of number of years of experience as a healthcare provider and experience with patients who were dying). The authors clearly indicate the hypotheses (or “hunches”) that stimulated their explorations of particular types of participants. Recruitment was done through letters of invitation, with a 60% response rate, which means the researchers would have sent out approximately 25 letters of invitation. The type of providers who decided not to participate was unclear from the report.

**Were the data analysed appropriately?**

All participants were interviewed once, and 3 providers were interviewed a second time. Initial interviews were done using open ended questions and lasted 60–90 minutes. Later interviews lasted 30–60 minutes as the questions became more focused. The authors included a table in their article that provided examples of changes in interview questions for participants 1–5, 6–10, and 11–15.

Although Norton and Bowers note that fieldwork was part of “member checking”, they did not fully describe the inclusion of a participant observation component in their research. When the grounded theory method was initially developed by Glaser and Strauss\(^1\), they included participant observation and interviews as data collection methods. At this time, most grounded theory studies only use interviews for data generation.

The authors did not incorporate an examination of records. A chart review might have illuminated what providers wrote about patient and family treatment choices and providers’ documentations of their attempts to influence those choices. The omission of such data does not weaken the study, but might have offered additional perspectives on the research question.

The study was done in a mid-size mid-western city in the US. Participating providers were recruited from home health and family practice, oncology practice, and intensive care units. The study was published in 2001, and the research was likely done approximately 2–5 years before that date. Although the authors did not clearly indicate the date of the study, a quick look at the references reveals that Norton completed her dissertation in 1999, and these data were collected during her dissertation.

Interviews were audiorecorded, transcribed verbatim, and checked for accuracy before data were entered into a computer qualitative data management system. Norton and Bowers used QSR NUD*IST 4 to assist in qualitative data management. Other procedures used to enhance the credibility of the findings reveal the authors’ attention to the analysis process. As principal investigator, Norton wrote that she was engaged in the collection and analysis of data for a period of 22 months, during which time she met weekly with a multidisciplinary grounded theory dimensional analysis group. Members of this group would have offered critique and commentary of ongoing analysis based on their disciplinary perspectives, thus enlarging the group’s breadth of views. This group’s work may have led to increased understanding of patient and family treatment choices and providers’ interpretations of incorporating the strategies, the authors noted how most strategies were used for more than one purpose. In presenting the number of strategies, the varying purposes of enacting a given strategy, and different interpretations of incorporating the strategies, the authors showed respect for the participants and were true to their purpose of examining the various ways that providers worked with patients and families at the end of life.

Norton and Bowers situated their findings within the literature they reviewed as background, using statements such as “Consistent with the findings from previous research studies...” and then listing those studies. Readers will sometimes find phrases such as “the results extend what was found by” and “in contrast to the findings of (reference), the results of this study suggest...”. The referent for such phrases will logically be found in the background section of the article.

Norton and Bowers consistently discussed decision making, treatment preferences, choices, and healthcare providers’ interactions with patients and families, all of which were areas explored in the results section. The authors clearly articulated
that future research is needed to explore patients’ and family members’ understandings of their conditions and decision making.

**What meaning and relevance does the study have for my practice?**

The study met Thorne’s 5 criteria for appraising the disciplinary relevance and usefulness of a study. Norton and Bowers clearly articulated the need for this research. Given that nurses and other healthcare professionals interact with patients and families as they make end-of-life treatment decisions, the topic is relevant to healthcare disciplines. The description of strategies that providers used in shifting patients and families from curative to palliative treatment decisions is illuminating for healthcare professionals who work in palliative care settings as well as for providers who work with patients and families around other important life decisions. The study situates itself within the historical context of advances in technology, complex end-of-life decisions, patients’ rights to self-determination, and advance directives. The authors concluded the article noting that only providers’ perspectives were presented and, even within that unique group, there was no one “right” or consistent way that providers engaged with patients and families. They also rightly point out what their study did not explore.

**Does the study help me understand the context of my practice? and Does the study enhance my knowledge about my practice?**

Norton and Bowers provided an adequate description of the context and setting of their study. The findings can sensitize providers to some of the implicit and unspoken ideas they may have and enact as they work with patients and families at the end of life. Framing their efforts as “work” legitimates the energy and time expended by providers. Additionally, the findings suggest the potential for exploring providers’ strategies for shifting patients’ and families’ treatment-related decisions in other contexts unrelated to palliative care and end of life.

**Resolution of the scenario**

After appraising the article, you return to your next team meeting to lead a discussion on counselling patients and family members on appropriate end-of-life treatment choices. You point out that patient, family, and provider decisions have been individually explored in various contexts. Limited research, however, has focused on understanding the intersection of patient, family member, and provider decision making about decisions for end of life or other treatments. You observe that even though the study by Norton and Bowers has important information about how providers used various strategies to shift patients’ treatment decisions, there was no consistent picture of how this was done or even if it should be done. You note that the researchers pointed out that some people might interpret providers’ use of strategies as paternalistic and possibly coercive. Given that this was a preliminary study, you caution your colleagues to avoid implementing the strategies in such a way as to influence the treatment decisions of patients and family members. Rather, you emphasise that one of the finer points of clinical applicability of this study is that of sensitising providers to ways that they may consciously or unconsciously act to influence patients’ and family members’ treatment decisions. You recommend that this topic be explored further on your unit.