Evaluative Criteria for Qualitative Research in Health Care: Controversies and Recommendations

Deborah J. Cohen, PhD
Benjamin F. Crabtree, PhD
Department of Family Medicine, Research Division, University of Medicine and Dentistry, Robert Wood Johnson Medical School, Somerset, New Jersey

ABSTRACT

PURPOSE We wanted to review and synthesize published criteria for good qualitative research and develop a cogent set of evaluative criteria.

METHODS We identified published journal articles discussing criteria for rigorous research using standard search strategies then examined reference sections of relevant journal articles to identify books and book chapters on this topic. A cross-publication content analysis allowed us to identify criteria and understand the beliefs that shape them.

RESULTS Seven criteria for good qualitative research emerged: (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. General agreement was observed across publications on the first 4 quality dimensions. On the last 3, important divergent perspectives were observed in how these criteria should be applied to qualitative research, with differences based on the paradigm embraced by the authors.

CONCLUSION Qualitative research is not a unified field. Most manuscript and grant reviewers are not qualitative experts and are likely to embrace a generic set of criteria rather than those relevant to the particular qualitative approach proposed or reported. Reviewers and researchers need to be aware of this tendency and educate health care researchers about the criteria appropriate for evaluating qualitative research from within the theoretical and methodological framework from which it emerges.


INTRODUCTION

Until the 1960s, the scientific method—which involves hypothesis testing through controlled experimentation—was the predominant approach to research in the natural, physical, and social sciences. In the social sciences, proponents of qualitative research argued that the scientific method was not an appropriate model for studying people (eg, Cicourel,1 Schutz,2,3 and Garfinkel4), and such methods as observation and interviewing would lead to a better understanding of social life in its naturally occurring, uncontrolled form. Biomedical and clinical research, with deep historical roots in quantitative methods, particularly observational epidemiology5 and clinical trials,6 was on the periphery of this debate. It was not until the late 1960s and 1970s that anthropologists and sociologists began introducing qualitative research methods into the health care field.4,7-14

Since that time, qualitative research methods have been increasingly used in clinical and health care research. Today, both journals (eg, Qualita-
tive Health Research) and books are dedicated to qualitative methods in health care, as well as specific information on focus groups, qualitative content analysis, observation and ethnography, interviewing, studying stories, and conversation, doing case study, action research, and a range of mixed-methods designs.

As part of a project to evaluate health care improvements, we identified a need to help health care researchers, particularly those with limited experience in qualitative research, evaluate and understand qualitative methodologies. Our goals were to review and synthesize published criteria for “good” qualitative research and develop a cogent set of evaluative criteria that would be helpful to researchers, reviewers, editors, and funding agencies. In what follows, we identify the standards of good qualitative research articulated in the health care literature and describe the lessons we learned as part of this process.

**METHODS**

A series of database searches were conducted to identify published journal articles, books, and book chapters offering criteria for evaluating and identifying rigorous qualitative research.

**Data Collection and Management**

With the assistance of a librarian, a search was conducted in December 2005 using the Institute for Science (ISI) Web of Science database, which indexes a wide range of journals and publications from 1980 to the present. Supplemental Appendix 1, available online-only at http://www.annfammed.org/cgi/content/full/6/4/331/DC1, describes our search strategy. This search yielded a preliminary database of 4,499 publications. Citation information, abstracts, and the number of times the article was cited by other articles were exported to a Microsoft Excel file and an Endnote database.

After manually reviewing the Excel database, we found and removed a large number of irrelevant publications in the physical and environmental sciences (eg, forestry, observational studies of crystals), and further sorted the remaining publications to identify publications in health care. Among this subset, we read abstracts and further sorted publications into (1) publications about qualitative methods, and (2) original research using qualitative methods. For the purposes of this analysis, we reviewed in detail only publications in the first category. We read each publication in this group and further subdivided the group into publications that (1) articulated criteria for evaluating qualitative research, (2) addressed techniques for doing a particular qualitative method (eg, interviewing, focus groups), or (3) described a qualitative research strategy (eg, sampling, analysis). Subsequent analyses focused on the first category; however, among publications in the second category, a number of articles addressed the issue of quality in, for example, case study, interviewing, focus groups, observation, discourse, and narrative research that we excluded as outside the scope of our analysis.

Books and book chapters could not be searched in the same way because a database cataloging these materials did not exist. Additionally, few books on qualitative methods are written specifically for health care researchers, so we would not be able to determine whether a book was or was not contributing to the discourse in this field. To overcome these challenges, we used a snowball technique, identifying and examining books and book chapters cited in the journal articles retrieved. Through this process, a number of additional relevant journal articles were identified as frequently cited but published in non–health care or non-indexed journals (eg, online journals). These articles were included in our analysis.

**Analysis**

We read journal articles and book chapters and prepared notes recording the evaluative criteria that author(s) posited and the world view or belief system in which criteria were embedded, if available. When criteria were attributed to another work, this information was noted. Books were reviewed and analyzed differently. We read an introductory chapter or two to understand the authors’ beliefs about research and prepared summary notes. Because most books contained a section discussing evaluative criteria, we identified and read this section, and prepared notes in the manner described above for journal articles and book chapters.

An early observation was that not all publications offered explicit criteria. Publications offering explicit evaluative criteria were treated as a group. Publications by the same author were analyzed and determined to be sufficiently similar to cluster. We examined evaluative criteria across publications, listing similar criteria in thematic clusters (eg, importance of research, conducting ethically sound research), identifying the central principle or theme of the cluster, and reviewing and refining clusters. Publications that discussed evaluative criteria for qualitative research but did not offer explicit criteria were analyzed separately.
Preliminary findings were synthesized into a Web site for the Robert Wood Johnson Foundation (http://www.qualres.org). This Web site was reviewed by Mary Dixon-Woods, PhD, a health care researcher with extensive expertise in qualitative research, whose feedback regarding the implications of endorsing or positing a unified set of evaluative criteria encouraged our reflection and influenced this report.

**RESULTS**

We identified 29 journal articles and 16 books or book chapters that offered explicit criteria for evaluating the quality of qualitative research. Supplemental Appendix 2, available online-only at http://www.annfammed.org/cgi/content/full/6/4/331/DC1, contains a table listing citation information and criteria posited in these works. An additional 29 publications were identified that did not offer explicit criteria but informed discourse on this topic and our analysis.

Seven evaluative criteria were identified: (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 verifiability or reliability. There was general agreement observed across publications on the first 4 quality dimensions, however, on the last 3 criteria, disagreement was observed in how the concepts of researcher bias, validity, and reliability should be applied to qualitative research. Differences in perspectives were grounded in paradigm debates regarding the nature of knowledge and reality, with some arguing from an interpretivist perspective and others from a more pragmatic realist perspective. Three major paradigms and their implications are described in Table 1.

**Fundamental Criteria**

It was widely agreed that qualitative research should be ethical, be important, be clearly and coherently articulated, and use appropriate and rigorous methods. Conducting ethically sound research involved carrying out research in a way that was respectful, humane, and honest, and that embodied the values of empathy, collaboration, and service. Research was considered important when it was pragmatically and theoretically useful and advanced the current knowledge base. Clarity and coherence of the research report were criteria emphasizing that the report itself should be concise and provide a clear and adequate description of the research question, background and contextual material, study design (eg, study participants, how they were chosen, how data are collected and analyzed), and rationale for methodological choices. Description of the data should be unexaggerated, and the relationship between data and interpretation should be understandable.

**Researcher Bias**

The majority of publications discussed issues of researcher bias, recognizing researchers’ preconceptions, motivations, and ways of seeing shape the qualitative research process. (It should be noted there is ample evidence to suggest researcher motivations and preconceptions shape all research.) One perspective (interpretivist) viewed researcher subjectivity as “something used actively and creatively through the research process” rather than as a problem of bias. A hallmark of good research was understanding and reporting relevant preconceptions through reflexive processing (ie, reflective journal-keeping). A second perspective (realist) viewed researcher bias as a problem affecting the trustworthiness, truthfulness, or validity of the account. In addition to understanding researchers’ motivations and preconceptions, value and rigor were enhanced by

---

**Table 1. Common Paradigms in Health Care Research**

<table>
<thead>
<tr>
<th>Paradigm</th>
<th>Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positivism</strong></td>
<td>There is a real world of objects apart from people</td>
</tr>
<tr>
<td>Researchers can know this reality and use symbols to accurately describe, represent and explain this reality</td>
<td></td>
</tr>
<tr>
<td>Researchers can compare their claims against this objective reality. This allows for prediction, control, and empirical verification</td>
<td></td>
</tr>
<tr>
<td><strong>Realism</strong></td>
<td>There are real-world objects apart from people</td>
</tr>
<tr>
<td>Researchers can only know reality from their perspective of it</td>
<td></td>
</tr>
<tr>
<td>We cannot separate ourselves from what we know; however, objectivity is an ideal researchers strive for through careful sampling and specific techniques</td>
<td></td>
</tr>
<tr>
<td>It is possible to evaluate the extent to which objectivity or truth is attained</td>
<td></td>
</tr>
<tr>
<td><strong>Interpretivism</strong></td>
<td>Reality as we know it is constructed intersubjectively</td>
</tr>
<tr>
<td>Meaning and understanding are developed socially and experientially</td>
<td></td>
</tr>
<tr>
<td>We cannot separate ourselves from what we know. Who we are and how we understand the world are linked</td>
<td></td>
</tr>
<tr>
<td>Researchers’ values are inherent in all phases of research. Truth is negotiated through dialogue</td>
<td></td>
</tr>
<tr>
<td>Findings or knowledge claims are created as an investigation proceeds and emerge through dialogue and negotiations of meanings among community members (both scholars and the community at large)</td>
<td></td>
</tr>
<tr>
<td>All interpretations are located in a particular context, setting, and moment</td>
<td></td>
</tr>
</tbody>
</table>

* References 26, 69, 70, 73, 77, 80, 94, 95, 98, 106.
† References 19, 26, 69, 70, 73, 75, 77, 84, 85, 87, 95, 107.
‡ References 19, 69, 70, 72, 73, 77, 80-82, 87, 94, 103, 105.
controlling bias through techniques to verify and confirm findings, as discussed in more detail below.\textsuperscript{5} Thus, whereas all publications agreed that researcher bias was an important consideration, the approach for managing bias was quite different depending on the paradigm grounding the work.

Validity

A number of publications framed the concept of validity in the context of quantitative research, where it typically refers to the “best available approximation to the truth or falsity of propositions.”\textsuperscript{142} Internal validity refers to truth about claims made regarding the relationship between 2 variables. External validity refers to the extent to which we can generalize findings. Across publications, different ideas emerged.

Understanding the concept of validity requires understanding beliefs about the nature of reality. One may believe that there can be multiple ways of understanding social life and reality, even multiple realities. This view of reality emerges from an interpretivist perspective. Hallmarks of high-quality qualitative research include producing a rich, substantive account with strong evidence for inferences and conclusions and then reporting the lived experiences of those observed and their perspectives on social reality, while recognizing that these could be multiple and complex and that the researcher is intertwined in the portrayal of this experience. The goal is understanding and providing a meaningful account of the complex perspectives and realities studied.\textsuperscript{1}

In contrast, research may be based on the belief that there is one reality that can be observed, and this reality is knowable through the process of research, albeit sometimes imperfectly. This perspective is typically associated with a positivist paradigm that underlies quantitative research, but also with the realist paradigm found in some qualitative research. Qualitative research based on this view tends to use alternative terms for validity (eg, adequacy, trustworthiness, accuracy, credibility) and emphasizes striving for truth through the qualitative research process, for example, by having outside auditors or research participants validate findings. An important dimension of good qualitative research, therefore, is plausibility and accuracy.\textsuperscript{1}

Verification or Reliability

Divergent perspectives were observed on the appropriateness of applying the concept of verifiability or reliability when evaluating qualitative research. As is validity, this concept is rooted in quantitative and experimental methods and refers to the extent to which measures and experimental treatments are standardized and controlled to reduce error and decrease the chance of obtaining differences.\textsuperscript{142} Two distinct approaches to evaluating the reliability of qualitative research were articulated. In the first, verification was a process negotiated between researchers and readers, where researchers were responsible for reporting information (eg, data excerpts, how the researcher dealt with tacit knowledge, information about the interpretive process) so readers could discern for themselves the patterns identified and verify the data, its analysis and interpretation.\textsuperscript{5} This interpretivist perspective contrasts with the second, realist, perspective. Rather than leaving the auditing and confirming role to the reader, steps to establish dependability should be built into the research process to repeat and affirm researchers’ observations. In some cases, special techniques, such as member checking, peer review, debriefing, and external audits to achieve reliability, are recommended and posited as hallmarks of quality in qualitative research.\textsuperscript{11} In Table 2 we provide a brief description of these techniques.

| Table 2. Verification Techniques Used in Qualitative Research |
|---|---|
| Technique | Definition |
| Triangulation | Using multiple data sources in an investigation to produce understanding |
| Peer review/ debriefing | The “process of exposing oneself to a disinterested peer in a manner paralleling an analytical session and for the purpose of exploring aspects of the inquiry that might otherwise remain only implicit within the inquirer’s mind” |
| External audits/ auditing | Auditing involves having a researcher not involved in the research process examine both the process and product of the research study. The purpose is to evaluate the accuracy and evaluate whether the findings, interpretations, and conclusions are supported by the data |
| Member checking | Data, analytic categories, interpretations, and conclusions are tested with members of those groups from whom the data were originally obtained. This can be done both formally and informally, as opportunities for member checks may arise during the normal course of observation and conversation |

Perspectives on the Value of Criteria

Health care researchers also discuss the usefulness of evaluative criteria. We observed 3 perspectives on the utility of having unified criteria for assessing qualitative research.

One perspective recognized the importance of validity and reliability as criteria for evaluating qualitative research.\textsuperscript{102,103} Morse et al make the case that without validity and reliability, qualitative research risks being
seen as nonscientific and lacking rigor.88,125 Their argument is compelling and suggests reliability and validity should not be evaluated at the end of the project, but should be goals that shape the entire research process, influencing study design, data collection, and analysis choices. A second approach is to view the criteria of validity and reliability as inappropriate for qualitative research, and argue for the development of alternative criteria relevant for assessing qualitative research.23

This position is commonly based on the premise that the theoretical and methodological beliefs informing quantitative research (from whence the criteria of reliability and validity come) are not the same as the methodological and theoretical beliefs informing qualitative research and are, therefore, inappropriate.136 Cogent criteria for evaluating qualitative research are needed. Without well-defined, agreed-upon, and appropriate standards, qualitative research risks being evaluated by quantitative standards, which can lead to assimilation, preferences for qualitative research that are most compatible with quantitative standards, and rejection of more radical methods that do not conform to quantitative criteria.82 From this perspective emerged a number of alternative criteria for evaluating qualitative research.

Alternative criteria have been open to criticism. We observed such criticism in publications challenging the recommendation that qualitative research using such techniques as member checking, multiple coding, external audits, and triangulation is more reliable, valid, and of better quality.72,82,90,91,112,127,143 Authors challenging this recommendation show how techniques such as member checking can be problematic. For example, it does not make sense to ask study participants to check or verify audio-recorded transcribed data. In other situations, study participants asked to check or verify data may not recall what they said or did. Even when study participants recall their responses, there are a number of factors that may account for discrepancies between what participants recall and the researcher’s data and preliminary findings. For instance, the purpose of data analysis is to organize individual statements into themes that produce new, higher-order insights. Individual contributions may not be recognizable to participants, and higher-order insights might not make sense.82 Similar issues have been articulated about the peer-review and auditing processes27,143 and some uses of triangulation.130 Thus, alternative criteria for evaluating qualitative research have been posited and criticized on the grounds that such criteria (1) cannot be applied in a formulaic manner, (2) do not necessarily lead to higher-quality research, particularly if these techniques are poorly implemented, and (3) foster the false expectation among evaluators of research that use of one or more of these techniques in a study is a mark of higher quality.72,81,90,91,112,123,127

A third approach suggests the search for a cogent set of evaluative criteria for qualitative research is misguided. The field of qualitative research is broad and diverse, not lending itself to evaluation by one set of criteria. Instead, researchers need to recognize each study is unique in its theoretical positioning and approach, and different evaluative criteria are needed. To fully understand the scientific quality of qualitative research sometimes requires a deep understanding of the theoretical foundation and the science of the approach. Thus, evaluating the scientific rigor of qualitative research requires learning, understanding, and using appropriate evaluative criteria.123,124,135,137

**DISCUSSION**

There are a number of limitations of this analysis to be acknowledged. First, although we conducted a comprehensive literature review, it is always possible for publications to be missed, particularly with our identification of books and book chapters, which relied on a snowball technique. In addition, relying on publications and works cited within publications to understand the dialogue about rigor in qualitative methods is imperfect. Although these discussions manifest in the literature, they also arise at conferences, grant review sessions, and hallway conversations. One’s views are open to revision (cf, Lincoln’s103,144), and relationships with editors and others shape our ideas and whom we cite. In this analysis, we cannot begin to understand these influences.

Our perspectives affect this report. Both authors received doctoral training in qualitative methods in social science disciplines (sociology/communication and anthropology) and have assimilated these values into health care as reviewers, editors, and active participants in qualitative health care studies. Our training shapes our beliefs, so we feel most aligned with interpretivism. This grounding influences how we see qualitative research, as well as the perspectives and voices we examine in this analysis. We have been exposed to a wide range of theoretical and methodological approaches for doing qualitative research, which may make us more inclined to notice the generic character of evaluative criteria emerging in the health care community and take note of the potential costs of this approach.

In addition, we use 3 common paradigms—interpretivism, realism, and positivism—in our analysis. It is important to understand that paradigms and debates about paradigms are political and used to argue for...
credibility and resources in the research community. In this process, underlying views about the nature of knowledge and reality have been simplified, sometimes even dichotomized (interpretivism vs positivism). We recognize our use of these paradigms as an oversimplification and limitation of our work, but one that is appropriate if only because these categories are so widely used in the works we analyze.

Our analysis reveals some common ground has been negotiated with regard to establishing criteria for rigorous qualitative research. It is important to notice that the criteria that have been widely accepted—carrying out ethical research and important research, preparing a clear and coherent research report, and using appropriate and rigorous methods—are applicable to all research. Divergent perspectives were observed in the field with regard to 3 criteria: researcher bias, validity, and verification or reliability. These criteria are more heavily influenced by quantitative and experimental approaches and, not surprisingly, have met with resistance. To understand the implications of these influences, our analysis suggests the utility of examining how these criteria are embedded in beliefs about the nature of knowledge and reality.

Central to the interpretivist paradigm, which historically grounds most qualitative traditions, is the assumption that realities are multiple, fluid, and co-constructed, and knowledge is taken to be negotiated between the observer and participants. From this framework emerge evaluative criteria valuing research that illuminates subjective meanings and understands and articulates multiple ways of seeing a phenomenon. Rich substance and content, clear delineation of the research process, evidence of immersion and self-reflection, and demonstration of the researcher’s way of knowing, particularly with regard to tacit knowledge, are essential features of high-quality research.

In contrast, fundamental to a positivist paradigm, which historically grounds most quantitative approaches, is the assumption that there is a single objective reality and the presumption that this reality is knowable. The realist paradigm softens this belief by suggesting knowledge of reality is always imperfect. Within the realist framework the goal of qualitative research is to strive for attaining truth, and good research is credible, confirmable, dependable, and transferable. Thus, rigorous qualitative research requires more than prolonged engagement, persistent observation, thick description, and negative case analysis, but it should use such techniques as triangulation, external auditing, and member checking to promote attainment of truth or validity through the process of verifying findings.

One reason for the centrality of the realist paradigm in health care research may be its ability to assimilate the values, beliefs, and criteria for rigorous research that emerge from the positivist paradigm. In a community that values biomedical bench research, sees the randomized controlled trial as a reference standard, holds a belief in an objective reality, and values research that is reliable, valid, and generalizable (typically positivist ideals), it is not surprising that realist views with regard to qualitative research have found favor. Unlike interpretivism, realism adopts a philosophy of science not at odds with the commonly held ideals of positivism. By maintaining a belief in an objective reality and posting truth as an ideal qualitative researchers should strive for, realists have succeeded at positioning the qualitative research enterprise as one that can produce research which is valid, reliable, and generalizable, and therefore, of value and import equal to quantitative biomedical research.

Although qualitative research emerging from a realist paradigm may have successfully assimilated into the clinical research community (as it has in other disciplines), it may be at a cost. Qualitative approaches most compatible with traditional values of quantitative research may be most likely to be accepted (published and funded). More radical methods (eg, feminist standpoint research, critical postmodern research), which can make innovative contributions to the field, may be marginalized because they do not fit the evaluative criteria that have emerged in the health care community. In addition, doing rigorous qualitative research in the way realists prescribe involves using a number of techniques that may foster the appearance of validity and reliability, but can be problematic if inappropriately applied.*

The search for a single set of criteria for good qualitative research is grounded in the assumption that qualitative research is a unified field. Qualitative research is grounded in a range of theoretical frameworks and uses a variety of methodological approaches to guide data collection and analysis. Because most manuscript and grant reviewers are not qualitative experts, they are likely to embrace a generic set of criteria. Reviewers and researchers need to be aware of the 7 criteria for good qualitative research, but also they need to be aware that applying the same standards across all qualitative research is inappropriate. Helping reviewers understand how an unfamiliar qualitative approach should be executed and standards for evaluating quality are essential, because reviewers, even qualitative experts, might not be well-versed in the particular qualitative method being used or proposed. Panel organizers and editors need to recognize

* References 72, 81, 90, 91, 112, 123, 127, 145.
that a qualitative expert may have only a very narrow range of expertise. Moreover, some researchers may be so entrenched in the dogma of their own approach that they are unable to value qualitative methods dissimilar from their own. This type of ax grinding harms not only the efforts of qualitative researchers, but the field more generally.

Future work needs to focus on educating health care researchers about the criteria for evaluating qualitative research from within the appropriate theoretical and methodological framework. Although the ideas posited here suggest there may be a connection between how quality is defined and the kind of work published or funded, this assumption is worthy of empirical examination. In addition, the field needs to reflect on the value of qualitative health care research and consider whether we have the space and models for adequately reporting interpretive research in our medical journals.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/cgi/content/full/6/4/331.

Key words: Qualitative research; methodology; health care research

Submitted April 15, 2007; submitted revised November 7, 2007; accepted November 14, 2007.

Funding support: Preparation of this report was supported by a grant from the Robert Wood Johnson Foundation (#053512).

Acknowledgments: We are indebted to Mary Dixon-Woods, PhD, for her insightful comments on earlier versions of this work.

References


118. Devers KJ. How will we know “good” qualitative research when we see it? Beginning the dialogue in health services research. Health Serv Res. 1999;34(5 Pt 2):1153-1188.


133. Poses RM, Levitt NJ. Qualitative research in medicine. Antirealism is an excuse for sloppy work. BMJ. 2000;320(7251):1729-1730.


