Guidelines for evaluating the levels of evidence based on quantitative research are well established. However, the same cannot be said for the evaluation of qualitative research. This article discusses a process members of an evidence-based clinical practice guideline development team with the Association of Women’s Health, Obstetric and Neonatal Nurses used to create a scoring system to determine the strength of qualitative research evidence. A brief history of evidence-based clinical practice guideline development is provided, followed by discussion of the development of the Nursing Management of the Second Stage of Labor evidence-based clinical practice guideline. The development of the qualitative scoring system is explicated, and implications for nursing are proposed. JOGNN, 31, 708–714; 2002. DOI: 10.1177/0884217502239216

Keywords: Evaluation of evidence—Evidence-based practice—Guideline development—Levels of evidence—Qualitative evidence—Qualitative research

Accepted: December 2001

The nursing profession’s recognition of the importance of evidence-based practice is well documented (Estabrooks, 1998; Gennaro, 1994; Hodnett, Kaufman, O’Brien-Pallas, Chipman, & Watson-MacDonell, 1996; Mayberry & Strange, 1997; Morin et al., 1999). Furthermore, much has been written about how to evaluate the evidence, particularly relative to quantitative studies (Carlson, Kruse, & Rouse, 1999; Cooke, 1996; Haughey, 1994; Mitchell, 1999; Mulhall, Alexander, & le May, 1998; Simpson & Knox, 1999). However, much of what nursing does and what nurses investigate is qualitative in nature (Beck, 1993; Morse & Field, 1995; Sandelowski, 1997; Streubert & Carpenter, 1995). Thus, evaluation of research literature should include both qualitative and quantitative reports (Ford-Gilboe, Campbell, & Berman, 1995; Goode, 2000). The evaluation, however, requires different sets of criteria to determine the merits of either type of report (Green & Britten, 1998; Popay, Rogers, & Williams, 1998). The purpose of this article is to discuss the process by which members of an Association of Women’s Health, Obstetric and Neonatal Nurses (AWHONN) evidence-based clinical practice guideline development team created a scoring system, based on specific criteria, for the evaluation of qualitative studies. The scoring system was designed to place qualitative studies within specific levels of evidence.

History of Evidence-Based Clinical Practice Guideline Development

With more than 2 million health-related manuscripts published in more than 20,000 health care journals every year, a mechanism was required to aid the health care provider in keeping abreast of this information explosion (Farquhar & Vandekerckhove, 1996). Archie Cochrane (1909–1988), a British obstetrician, gynecologist, epidemiologist, and researcher, suggested that although the randomized controlled trial was the research method most likely to yield reliable results, it was impossible for a clinician to easily access all of the latest studies potentially affecting practice in each clinical special-
ty. Many early attempts to summarize or synthesize clinical research were of poor quality, however, leading to conclusions that were biased or incorrect (Antman, Lau, & Kupelnick, 1992). Cochrane was one of the first health care providers to question the level of evidence available to assist in making medical decisions. In response to the increase in published information and the potential to disseminate misinformation, Cochrane developed a process by which research was systematically and objectively reviewed for methodologic rigor and clinical feasibility (Cochrane, 1979). He made the recommendation to the U.K. National Health Service that only health care practices that have been studied, ideally as randomized controlled trials, and deemed beneficial and cost effective be incorporated into care standards of the National Health Service in that country (Cochrane, 1972).

In 1992, the U.K. Cochrane Center, a network of institutions and individuals, was established to provide a reliable summary of information on selected health care problems addressed in approximately 1 million published studies. Only 1 year later, the group was expanded and the Cochrane Collaboration was formed. Since 1993, the international component of the Cochrane Collaboration has evolved and now includes centers in Australia, Canada, Italy, Scandinavia, The Netherlands, France, and the United States (Farquhar & Vandekerckhove, 1996).

Today, the Cochrane Collaboration maintains a registry of trials, provides organizational support, and coordinates groups to avoid duplication of efforts on a specific topic. When a review has been completed, the authors may choose to have it included in the Cochrane Database of Systematic Reviews. This database, along with the Database of Abstracts of Reviews of Effectiveness, the Cochrane Controlled Trial Register, and the Cochrane Review Methodology Database, can be accessed online via the Cochrane Library (Chalmers, Dickersin, & Chalmers, 1992). The Web site is listed in Table 1 along with other Web resources that support evidence-based guideline development. Thus, these organizations have made concerted efforts to facilitate the use of research in practice. Furthermore, because of the increased accountability of health care institutions and providers, the rising cost of health care, and the difficulty in measuring quality, health care organizations seek to identify evidence-based processes to address high-quality, cost-efficient health care modalities (Zinberg, 1997).

As the standard bearer for the nursing care of women and newborns, AWHONN has demonstrated its commitment to advancing research-based practice. This commitment is demonstrated by the development of evidence-based clinical practice guidelines through the systematic review and scoring of the literature by an appointed team of AWHONN volunteers and by publication of the guideline.

Criteria have been established to facilitate the process of evidence-based guideline development (Sackett & Haynes, 1976). AWHONN has taken a comprehensive approach to evidence-based clinical practice guideline development based on the framework presented in the ANA Manual to Develop Guidelines (Marek, 1995). Several approaches for evaluating evidence are included in this framework. The ANA manual models its process on that of the Agency for Health Quality Research, formerly the Agency for Health Care Policy and Research. The

| **TABLE 1**
| Internet Resources for Evidence-Based Guideline Development |
|---|---|
| **The Cochrane Collaboration** | http://www.cochrane.org/ |
| **BMJ Publication – Evidence-Based Nursing** | http://www.evidencebasednursing.com/ |
| **Sigma Theta Tau International—The Online Journal of Knowledge Synthesis for Nursing** | http://www.stti-web.iupui.edu |
| **Association of Women’s Health, Obstetric and Neonatal Nurses** | http://www.awhonn.org |
| **Health Web** | http://www.uic.edu/depts/lib/health/hw/ebhc |
| **Medical Matrix** | http://www.medmatrix.org |
| **Hardin Library** | http://www.lib.uiowa.edu/hardin/md/index/html |

The ability to evaluate the level of evidence of qualitative research has broad implications for nursing.

November/December 2002
manual emphasizes that research-based guidelines which address significant topics “are generalizable to the expectations of the scientific and clinical communities both in nursing and the larger health care community” (Marek, 1995, p. iii).

One approach, drawn from the Agency for Health Care Policy and Research’s Panel for the Prediction and Prevention of Pressure Ulcers in Adults, asks reviewers to rate the degree to which a report meets selected criteria on a scale of 0 (no evidence that criterion met) to 3 (good) (Marek, 1995). Criteria address the following: the problem statement, sampling, measurement, internal validity, external validity, construct validity, statistical conclusion validity, and justification for conclusions. Each research report receives a total score based on how well criteria were met. The report is then categorized based on the strength of evidence or type of research being reported and based on the U.S. Preventive Services Task Force (1996) criteria. Thus, category “I” indicates that the report was based on at least one well-designed randomized controlled trial whereas category “III” indicates the report was reflective of the opinion of respected authorities or based on clinical experience, descriptive studies, or reports of expert committees. Another approach (Brown, 1991) rates the design, selection, and specification of the sample; the specification of illness or condition; the description of educational condition; the construct definition of the outcome; and outcome measures.

Development of the Nursing Management of the Second Stage of Labor Evidence-Based Clinical Practice Guideline

AWHONN’s research-based practice program provides a systematic, participative approach to the design, implementation, and evaluation of evidence-based practice guidelines for the use of practitioners in clinical settings (AWHONN, 2000). AWHONN’s journey to guideline development for the management of the second stage of labor began in 1993 with a research utilization project titled Second Stage Labor Management (RU2) (Mayberry & Strange, 1997). The goals of the RU2 were (a) to assist in promoting the birth of the uncompromised fetus by minimizing negative maternal hemodynamic changes caused by inappropriate positions and pushing techniques and (b) to minimize maternal fatigue. The project was conducted over a 2-year period at 40 hospitals: 38 in the United States and 2 in Canada. The primary focus of the original RU2 project was promoting positioning and pushing technique alternatives.

Nurses and agencies who participated in RU2 received information regarding research related to second-stage labor management, the project guideline focusing on positioning and pushing techniques, and the rationale for the use of these methods of care delivery. Because second-stage labor management may overlap with the domains of other disciplines such as medicine and anesthesiology, these health care providers were also included in the development of the project guideline. The research utilization project provided valuable insight into the process of changing nursing practice based on available research.

The Process of Guideline Development

The results of the RU2 project led to a call for volunteers in late 1998 to serve on AWHONN’s Nursing Management of the Second Stage of Labor Evidence-Based Clinical Practice Guideline Development Team. Recommendations and findings from the project provided the basis for the development of this evidence-based clinical practice guideline. Throughout 1999 and 2000, team members participated in teleconferences; literature review, evaluation, and scoring; ongoing progress reports and process evaluation; and creation of the guideline.

Nurses, as do qualitative researchers, tend to approach situations from a worldview that is comprehensive and holistic, rather than reductionistic and deterministic.

Group consensus was employed to facilitate decisions related to the type and the publication years of the literature (multidisciplinary) reviewed. The search was conducted using both MEDLINE and CINAHL databases. The search was also limited to articles in English published between 1990 and 1999. Articles reporting the results of clinical trials, review articles, and reports of single studies also were reviewed and scored. Duplicate articles were eliminated.

The search was limited to articles in which any one of the terms labor, obstetrical nursing, labor complications, or delivery were present. In addition, these terms were combined with each of the following terms: social support, nurse-patient relationships, and helping behavior. Additional search terms included posture, bearing down, valsalva, or pushing. The last three words were searched within the text. Reports not specifically related to second stage were eliminated.

Each research report was evaluated using a scoring tool to rate the quality of each study based on the criteria described above. The criteria derived from the ANA Manual to Develop Guidelines (Marek, 1995) were selected because they were considered to be comprehensive and user-friendly. However, several relevant studies reviewed received low scores because they were qualitative. The team recognized that because the scoring tool was based
on criteria for evaluating quantitative literature, different criteria were required to evaluate the quality of qualitative evidence.

**Development of the Qualitative Scoring System**

The team reviewed literature on methods of evaluating qualitative studies (Cobb & Hagenmaster, 1987; Forbes et al., 1999; Greenhalgh & Taylor, 1997; Marshall & Rossman, 1995; Morse, Penrod, & Hupcey, 2000; Shields, 1997). The Cochrane handbook provided a useful section addressing the process of rating the quality of qualitative methods (Clarke & Oxman, 1999). Kearney (2001) suggested that the evaluation of qualitative studies should be extended beyond the rigor with which the evidence was developed to evaluation of the characteristics of the findings themselves.

Although criteria from all of the above methods were explicated, none met the needs of the team relative to generating a unique score that would adequately reflect the strength of qualitative evidence. Because the process for guideline development was in progress, team members were familiar with the tools for rating quantitative literature. The team agreed to identify criteria for evaluation of qualitative literature that could be adapted into a format similar to the existing quantitative rating tool. Consequently, the AWHONN staff facilitator and the team generated a scoring tool based on evaluative criteria for qualitative research proposed by Burns and Grove (1999).

**Archie Cochrane was one of the first health care providers to question the level of evidence available in making medical decisions.**

Categories described by Burns (1989) containing the criteria for rating were the following: descriptive vividness, methodological congruence, analytical preciseness, theoretical connectedness, and heuristic relevance.

Elements of these categories were judged by the team to be adaptable for the development of a scoring tool designed to facilitate accurate evaluation of qualitative studies. Consistent with the format for rating quantitative studies, two guideline development teams created a numeric scoring system of 0 to 3 to rate the qualitative studies. Thus, the qualitative research reports were rated using the criteria within each of the five categories described above. On this scale, 3 indicated that more than 75% of the criteria had been met, 2 that between 50% and 74% of the criteria had been met, and 1 that between 25% and 49% of the criteria had been met. A score of zero indicated that there was no evidence that criteria were met. The quality of evidence for each study was then determined using a designation of QI, QII, or QIII, based on evaluation of the total scores. For example, a quality of evidence rating of QI was consistent with a high total numeric score indicating that 75% to 100% of the total criteria in each of the five categories had been met. Therefore, the designation QI is indicative of a high-level, well-constructed, qualitative study. A detailed description of the scoring criteria, sample scoring sheet, and quality of evidence rating can be found in Tables 2 and 3.

**TABLE 2**

Instructions for Evaluating Qualitative Literature

- Read the qualitative study and score each of the categories listed using the quality rating scale of 0 through 3 described below.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DV</td>
<td>Descriptive Vividness</td>
</tr>
<tr>
<td>MC</td>
<td>Methodological Congruence</td>
</tr>
<tr>
<td></td>
<td>a. RD = Rigor in Documentation</td>
</tr>
<tr>
<td></td>
<td>b. PR = Procedural Rigor</td>
</tr>
<tr>
<td></td>
<td>c. ER = Ethical Rigor</td>
</tr>
<tr>
<td></td>
<td>d. C = Confirmability</td>
</tr>
<tr>
<td>AP</td>
<td>Analytical Preciseness</td>
</tr>
<tr>
<td>TC</td>
<td>Theoretical Connectedness</td>
</tr>
<tr>
<td>HR</td>
<td>Heuristic Relevance</td>
</tr>
<tr>
<td></td>
<td>a. IR = Intuitive Recognition</td>
</tr>
<tr>
<td></td>
<td>b. RBK = Relationship to Existing Body of Knowledge</td>
</tr>
<tr>
<td></td>
<td>c. A = Applicability</td>
</tr>
</tbody>
</table>

**SCORING SCALE**

- 3 = Good = 75%–100% criteria met
- 2 = Fair = 50%–74% criteria met
- 1 = Poor = 25%–49% criteria met
- 0 = No evidence that criteria met = < 25% criteria met

**FINAL QUALITY OF EVIDENCE RATING**

The quality of evidence rating was based on the total scores for each of the five categories described above. A quality of evidence rating for each qualitative study was assigned using the legend below:

- QI: Total score of 22.5–30 indicates that 75% to 100% of the total criteria were met.
- QII: Total score of 15–22.4 indicates that 50% to 74% of the total criteria were met.
- QIII: Total score of less than 15 indicates that less than 50% of the total criteria were met.

November/December 2002
TABLE 3
Detailed Explanation of Criteria

Category I: Descriptive Vividness
- Is essential descriptive information included?
- Is there clarity in the description of the study?
- Is there credibility in the description of the study?
- Is there adequate length of time spent at the site to gain the familiarity necessary for vivid description?
- Does the researcher validate findings with the study participants?
- Is the descriptive narrative written clearly? (vividly?)

Category II: Methodologic Congruence
A. Rigor in documentation
- Are all elements or steps of the study presented accurately and clearly?
  1. Introduction
     - Phenomenon is identified
     - Philosophical base of study is made explicit
     - Purpose and type of qualitative study is stated
     - Study questions or aims are identified
     - Assumptions are identified
  2. Literature review
  3. Statements of methods
     - Access to site, sample, and population
     - Researcher’s role and interview structure
  4. Data collection
  5. Data analysis
  6. Conclusions/findings
B. Procedural rigor
- Has the researcher asked the right questions? Does the researcher tap the participant’s experience versus her or his theoretical knowledge of the phenomenon?
- Did the researcher describe steps taken to ensure that the participant did not misrepresent herself or himself, or misinform the researcher?
- Did the researcher describe steps taken to deter the informant from substituting supposition about an event rather than recalling the actual experience?
- Did the researcher eliminate the potential for “elite bias” by placing equal weight on high-status or elite informant data and low-status or less articulate informant data?
- Did the researcher describe steps taken to avoid influence or distortion of the events observed by her or his presence? (like the Hawthorne effect)
  - Were sufficient data gathered?
  - Was sufficient time spent gathering data?
  - Were the approaches for gaining access to the site or participants appropriate?
  - Was the selection of participants appropriate?
C. Ethical rigor
- Were participants informed of their rights?
- Was informed consent obtained from the participants and documented?
- Were mechanisms developed and implemented to protect participants’ rights?

D. Confirmability (auditability)
- Was the description of the data collection process adequate?
- Were the records of the raw data sufficient to allow judgments to be made?
- Did the researcher describe the decision rules for arriving at ratings or judgments?
- Could other researchers arrive at similar conclusions after applying the decision rules to the data?
- Did the researcher record the nature of the decisions, the data on which they were based, and the reasoning that entered into the decisions?

Category III: Analytical Preciseness
- Did the interpretive theoretical statements correspond with the findings?
- Did the set of themes, categories, or theoretical statements depict or describe a whole picture?
- Can the hypotheses or propositions developed during the study be verified by data?
- Were the hypotheses or propositions presented in the research report?
- Are the study conclusions based on the data gathered?

Category IV: Theoretical Connectedness
- Are the theoretical concepts adequately defined and/or validated by data?
- Are the relationships among the concepts clearly expressed?
- Are the proposed relationships among the concepts validated by data?
- Does the theory developed during the study yield a comprehensive picture of the phenomenon under study?
- Is a conceptual framework or map derived from the data?
- Is there a clear connection made between the data and the (nursing) frameworks?

Category V: Heuristic Relevance
A. Intuitive recognition
- Is the phenomenon described well?
- Would other researchers recognize or be familiar with the phenomenon?
- Is the description of the phenomenon consistent with common meanings or experiences?
B. Relationship to existing body of knowledge
- Did the researcher examine the existing body of knowledge?
- Was the process studied related to nursing and health? (do we need this?)

C. Applicability
- Are the findings relevant to nursing practice?
- Are the findings important for the discipline of nursing?
- Can the findings contribute to theory development?

(continued)
eral use. Overall, the guideline development teams expressed satisfaction and a sense of achievement in creating a unique and practical method to evaluate qualitative literature.

**Implications for Nursing**

The ability to evaluate the level of evidence of qualitative research has broad implications for clinical nursing practice. In addition, evaluating levels of evidence of qualitative research can highlight the unique contribution that nursing makes relative to patient care that is easily distinguishable from the contributions of medicine and other health care professions. This assertion is based on the belief that nurses, as do qualitative researchers, tend to approach situations from a worldview that is comprehensive and holistic, rather than reductionistic and deterministic. Nurses, as do qualitative researchers, appreciate that there is no single reality. Rather, reality changes over time, is unique to each person, and is context dependent (Burns & Grove, 1999).

Although many nursing care activities have an anatomical and/or physiologic basis that can be evaluated through randomized controlled trials, much of the nurse’s role involves providing holistic, supportive, or informational care and advocacy. Outcomes of this care in a maternal newborn setting are often measured by qualitative descriptions of pregnancy, childbirth, or parenting experiences of satisfaction, feelings, and self-efficacy. Amassing a body of evidence that supports the uniqueness of nursing promotes the autonomy, distinctiveness, and importance of our discipline. However, as with quantitative studies, there is considerable variation in the rigor with which qualitative studies are conducted.

The development of a systematic process to evaluate qualitative methods will be beneficial for nursing. The qualitative approach is gaining popularity and acceptance as a valid research forum. Many new journals focus on qualitative methods, and established professional journals, historically aimed at quantitative reports, are publishing more qualitative studies (Polit & Hungler, 1999).

To capture the richness and depth of qualitative findings, it is important for the reader to develop a systematic process for evaluating the content of the study report. The user-friendly scoring criteria developed by the AWHONN Nursing Management of the Second Stage of Labor Evidence-Based Guideline Development Team can provide the reader with such information. This is of particular importance if the evaluation of qualitative research was not a component of the reader’s formal nursing education.

Qualitative research findings can provide important information when there is little quantitative information to support nursing practice. These studies may address an issue alone or in conjunction with quantitative approaches to enhance the theoretical basis and/or validity of study findings.

The key to using quantitative and qualitative research to support or change nursing practice hinges on the ability to discern well-constructed studies from those that are not. With the use of AWHONN’s criteria and instructions for evaluating qualitative literature, the nurse is one step closer to gaining a better understanding of qualitative methods and incorporating the findings of rigorous studies into practice, thereby enhancing the provision of holistic care.

**Acknowledgments**

The authors thank the AWHONN staff and the other members of the Second Stage of Labor Management Evidence-Based Practice Guideline Development Team, Susan Longacre, Michelle Murray, Jacquelyn Reid, Karen Trapani, and Susan Walsh, for their assistance in the development of the qualitative scoring instrument.

**REFERENCES**


Popay, J., Rogers, A., & Williams, G. (1998). Rationale and standards for the systematic review of qualitative literature in health services research. Qualitative Health Research, 8, 341-351.


Sandra Cesario is an assistant professor, College of Nursing, Texas Woman’s University, Houston.

Karen Morin is professor of nursing and interim director, School of Nursing, The Pennsylvania State University, University Park, PA.

Anne Santa-Donato is associate director, Childbearing and Newborn Programs, Association of Women’s Health, Obstetric and Neonatal Nurses, Washington, DC.

Address for correspondence: Sandra K. Cesario, RNC, PhD, 3511 Southdown Drive, Pearland, TX 77584-2367. Email: SCesario@twu.edu.