

THE DELPHI TECHNIQUE AS A METHOD FOR SELECTING CRITERIA TO EVALUATE NURSING CARE

Patricia Farrell • Kathleen Scherer

The identification of valid and reliable indicators of quality nursing care has been recognized as a priority for nursing research (Lindeman, 1975). In Canada, over the past decade, the response to this recognition has been the development of nursing standards (Manitoba Association of Registered Nurses, 1977; Chagnon, 1977; Alberta Association of Registered Nurses, 1981; College of Nurses of Ontario, 1976; Canadian Nurses Association, 1980).

As nurses operationalize the construct "quality nursing care," they are confronted with two tasks. The first task involves the description of the indicators of quality nursing care. The establishment of content validity of indicators has been lacking in the literature. Once the indicators have been described, the second task is their qualification. To date, most projects have focused on the first task. A common approach in these projects has been the selection of a small panel of clinical experts, which has been used to identify or to establish the face validity of the descriptors of quality care.

The project reported here was also concerned with the first task, that of describing indicators of quality nursing care. However, in selecting indicators and establishing their content validity, this project challenged the previous assumption that clinical experts could be viewed as representative of all practising nurses. Instead, the basic assumption was made that the opinions of all practising nurses were the most valid source of information about what constitutes quality nursing care. A second assumption was that these opinions varied widely and were dependent upon education, experience, and practice setting. In order to obtain a large sample of actively practising nurses and to minimize costs, a survey approach was used. The Delphi technique emerged from the review of survey methodologies as an approach that would elicit opinions and mold consensus.

Patricia Farrell, R.N., M.Sc.(A), Assistant Professor of Nursing, University of Manitoba at Winnipeg.

Kathleen Scherer, R.N., M.H.Sc., Nursing Consultant-Standards, Manitoba Association of Registered Nurses.

They are Co-principal Investigators for the Manitoba Nursing Standards Project.

LITERATURE REVIEW

The Delphi Technique

This review primarily will focus on the use of the Delphi technique as a method for selecting criteria to evaluate patient care. The Delphi technique is used to structure the communication process and to allow individuals to deal with complex problems as a group. While the Delphi questionnaire format is similar to that of other survey formats, there are important distinctions. The distinguishing features of the Delphi technique include (a) a series of questionnaire administrations (rounds), (b) provision of feedback to each respondent following each round, consisting of the individual's response to each item and the group's mean or average response, and (c) an iteration process in which individuals are asked to reconsider their former opinions in light of the above feedback.

The Delphi technique may be particularly useful in a content area where there is divergence of opinion, and where empirical data are lacking. With this technique, all respondents are anonymous, a characteristic which is particularly helpful in avoiding the effect of influential leaders. The anonymity provided in the method guards against the influence of quantity or strength of personality (Linstone & Turoff, 1975).

While there are numerous modifications of the Delphi technique, three categories of its use are relevant to this review. Weaver (1972) distinguished between the exploratory Delphi, used to develop projections of future events, and the normative Delphi, used to facilitate goal-formation and thus shape the future. A third category may be called reactive, as panel members are asked to react to prepared information rather than generate ideas or items. Bramwell and Hykawy (1974) and Lemieux-Charles (1980) in nursing, and Bender, Strack, Ebright and Von Haunalter (1969) in medicine used an exploratory Delphi technique to identify expected changes in their professions, which could provide a baseline for measurement of real change. The normative Delphi was used by Lindeman (1975) to explore priorities in clinical nursing research, by Oberst (1978) to determine priorities for research in cancer nursing and by Ventura and Waligora-Serafin (1981) to identify priorities in mental health nursing.

Literature concerned with the use of the Delphi technique in the evaluation of patient care includes Milholland's survey of 14 general surgeons to estimate human surgical and mortality rates associated with liver trauma (Milholland, Wheeler, & Heieck, 1973). Osborne and Thompson (1975) administered a Delphi survey to a panel of 452 pediatric experts for their decision on criteria to evaluate ambulatory child health care. Holmes used a modified Delphi technique with seven medical faculty members to generate performance-based outcome criteria for evaluation of family-centered primary care (Holmes, Kane, Ford, & Fowler, 1978). Inui, Hill and Leiby (1979) used the method with nine physicians to develop setting-specific standards of care for diagnosis and prevention of tuberculosis. Romm and Hulka (1979) examined the effect of the Delphi iteration procedure in developing explicit criteria for ambulatory care assessment in a study involving thirty-one practising physicians who reacted to disease-specific criteria. Another study used a Delphi procedure with academic medical clinicians to develop criteria for ambulatory care (Hastings, Sonneborn, Lee, Vick, & Sasmor, 1980). More recently the Alberta Hospital Association has chosen the Delphi technique to develop outcome criteria for the evaluation of home care services (Shields, Note 1).

In all of the literature reviewed, content validity of the criteria sets was not reported. Measurement theory (Anastasi, 1968) has specified content validation procedures which include: (a) a review of the literature to identify the behaviour domain to be measured, (b) consultation with subject matter experts to identify the content sampling of the behaviour domain, and (c) use of resulting content sampling to serve as the test specifications.

From the literature review the practice of nursing was conceptualized as a universe with four behaviour domains: research, education, administration, and clinical nursing or direct nursing care. Direct nursing care was selected as the behaviour domain for the development of standards. Three frameworks which further defined the behavioural domain to be measured were also identified. The first framework, the nursing process, was generally accepted as nursing's application of the scientific method and was practised widely. The nursing process had also been used as a basis for the evaluation of nursing care (Hegyvary, 1979). Donabedian's (1966) health care evaluation model of structure, process and outcome was retained. Williamson's (1978) outcome model was adapted to the nursing process. Additionally, Bloch's (1977) definitions of criteria and standards were incorporated, as follows:

A criterion is the value-free name of a variable believed or known to be a relevant indicator of the quality of patient care.

A standard is the desired and achievable level of performance corresponding with a criterion against which actual performance is compared.

In this manner, the first requirement for the establishment of content validity was satisfied.

The second requirement for the establishment of content validity was the identification of the content sampling of the behaviour domain. The Manitoba Association of Registered Nurses Standards Committee, through a series of twelve drafts, described items which were indicators of the content sampling of the behaviour domain (Scherer, Cameron, Farrell, Ramsay, & Vogt, 1982).

METHODOLOGY

This study was a collaborative project between the University of Manitoba School of Nursing and the Manitoba Association of Registered Nurses, conducted from October 1979 to May 1981.* The population of active practising nurses in Manitoba was sampled and supplemented with Canadian nurse educators and researchers and American nurse researchers. A Delphi questionnaire was designed and distributed. A follow-up of non-respondents was conducted. Results were analyzed and the quantitative and qualitative findings were used in the refinement of the second edition of the M.A.R.N. Standards of Nursing Care (1981).

Sampling Procedure

Four major strata were identified as representative of the nursing population: nurse administrators, nurse educators, nurse researchers, and nurses involved in direct nursing care. Operational definitions which would be compatible with existing data sources of provincial registered authorities were developed for each strata. Underrepresentation of Manitoba nurse educators and nurse researchers necessitated enlargement of the population to include Canadian nurse educators and researchers and American nurse researchers (Scherer, Cameron, Ramsay, Vogt, & Farrell, 1981). The population of 8980 nurses was enumerated. Proportional random sampling of this population resulted in 1411 nurses who were invited to participate in the Delphi survey. Of these, 662 indicated their agreement to participate by returning a sociodemographic profile.

* Scherer, K. & Cameron, C. A project to measure the quality of nursing practice in the province of Manitoba was funded by a grant from the M.A.R.N. Research Fund.

Questionnaire Design

In designing the Delphi questionnaire, one of the content validation tasks was to ascertain not only that the standards were 'desirable' but also that they were 'necessary.' Two questions were developed and a five-point Likert response scale was used. This scaling technique was used in an attempt to minimize the influence of the midpoint (centering bias) and yet encourage diverse expression of opinions.

STANDARDS OF NURSING CARE

Instructions: Read the first item of the standards of nursing care. Then circle the answer (for each of the questions), which best represents your opinion about the item. Continue on with the second and each remaining item.

QUESTION 1

In general, across all types of nursing settings, this item must be present and/or put into practice in order for good nursing care to be provided. Do you strongly disagree, moderately disagree, neither agree nor disagree, moderately agree or strongly agree?

QUESTION 2A

This item may or may not apply in your setting(s) for various reasons. However, the present standards indicate that each item *should* apply. Do you strongly disagree, moderately disagree, neither agree nor disagree, moderately agree or strongly agree?

Figure 1. Likert Scale in Delphi Questionnaire.

The Likert categories were scaled from one to five: (1) strongly disagree, (2) moderately disagree, (3) neither agree nor disagree, (4) moderately agree, and (5) strongly agree. The five discrete categories resulted in interval level data. An a priori decision was made to accept items with a mean response score greater than 4. Items with a mean of 3 or less were rejected. Following qualitative appraisal, items in which the mean was between 3 and 4 were retained if the standard deviation was within +1.

Beside each question, space was provided for written comments. The questionnaire was pretested with 16 nurses, and was also translated into French. The questionnaire was mailed to each respondent with stamped, self-addressed, return envelopes. Follow-up cards were sent to non-respondents four weeks later.

Data Analysis

Four hundred and seventy-two nurses (71%) completed round one of the Delphi questionnaire. The grand mean response to question 1 was 4.59 and to question 2, 4.5, both indicating moderate agreement. Responses clustered about the mean within one standard deviation. Only two items were rejected based on their mean of less than 3. When frequency distribution of the responses within each strata was undertaken, it was found that 10-30% of the nurse educators either moderately disagreed or registered a neutral vote when compared to the grand mean. In order to determine whether this strata would alter their responses when provided with feedback on the entire group's response, a second round of the Delphi was administered to the 171 nurse educators.

A computer-generated round two of the Delphi survey consisted of the following feedback: (a) the group's mean response to each item and (b) that individual's previous response to each item. Researchers were asked to reconsider their responses in light of the group's response and to complete the questionnaire again. Follow-up cards were mailed to non-respondents. An 84% response rate (141 nurses) was obtained on round two of the Delphi. While the grand mean for questions 1 and 2 remained greater than 4 (moderately agree), the frequency distribution indicated that 10% moderately disagreed and 22% neither agreed nor disagreed. While 10% represented no change in strongly held opinions (the moderately disagree category), the neutral category was shifted upward (8%) between rounds one and two of the Delphi.

Quantitative responses to both rounds were content analyzed, summarized and forwarded to all respondents (Scherer, Cameron, & Farrell, 1981). All qualitative responses were considered in the process of content validation of the M.A.R.N. Standards of Nursing Care.

DISCUSSION

Two major limitations of this study concerned the English-to-French translation and the process for identification of the population. Although the questionnaire was professionally translated into French, comprehension difficulties were experienced by some French respondents, necessitating a re-translation and re-administration. Thus, questions about reliability of the French responses may be raised. A back-translation procedure and pre-testing should be used to minimize this effect in a bilingual country. The second limitation concerned identification of the population. Since no national data base exists, six months were required to access existing data sources for

population identification. Where these were deficient, new data sources were constructed. The fact that full-time nurse researcher positions are rare in Canada meant that nurse educator and nurse researcher positions were not mutually exclusive. This overlap in the operational definition variable, 'employed full-time', led to over-representation of nurse educators in the sample. This over-representation reflects actual practice and at this time is impossible to avoid.

In this study the second round of the Delphi resulted in no change in negative group opinion (moderately disagree). While this analysis did not examine individual changes, it could be said that opinions which were strongly held were not influenced by the Delphi iteration procedure. However, the process of individual and group feedback did effect some positive shift to the moderately agree in those who previously held neutral opinions.

In general, the basic assumption that nurses hold opinions about what constitutes quality nursing care was supported by this study. While previous research used exclusive panels of clinical experts, this project was designed to view all practising nurses as having expertise. This approach was also intended to prevent the imposition of standards developed by a non-representative group upon the general nursing population. While no significant variation was found among the four strata, membership acceptance and utilization are stronger when all members participate in standards development.

The second assumption of the project, that opinions about quality nursing care vary widely, is unresolved. The first round of the Delphi produced an overall consensus about the standards and the second round did not alter that consensus. While the results may seem to support Romm and Hulka's (1979) conclusion that a second mailing does not make an appreciable difference in the selection of measurement criteria, it is questionable whether a consensus would have been obtained if the first unrefined draft of the questionnaire had been administered.

In the examination of traditional response biases, the range of responses and resulting group means indicated that centering biases had been successfully reduced. Alternatively, the participants' expectancy that a consensus was the objective may have encouraged socially desirable or acquiescent responses. While no sociodemographic differences were found between respondents and drop-outs, analysis of non-respondents (those who refused to participate) was not possible. Non-respondents may be distinctly different in their social

acquiescence/deviance response set and this warrants further investigation. Also, reverse ordering of the Likert scale could have reduced this response set, and order effects require further study. A heuristic question may exist in considering whether the results from this project represent a socially desirable response set or nurses' beliefs that clients deserve quality nursing care as defined in the Standards. Whether the care which nurses deliver and patients receive approximates the Standards is the ultimate question.

The sample of experts used in this study represented the spectrum of nurses engaged in nursing. This not only added to the validity of the process but could serve as a motivating factor in terms of commitment to the Standards. The generic data base which resulted from this content validity study is now being used for instrument development to measure criteria and standards — the next step in quality assurance.

CONCLUSION

In the area of measurement of quality of care, the nursing profession has two options. One would be to wait until we have sufficient empirical evidence from which to evaluate nursing care; the other is to generate standards by using the judgment of experts in the field. Since little empirical evidence exists in the area of what constitutes quality nursing care, this project used the second option. In this study, experts in the field comprised representatives of the entire nursing population in Manitoba and a selected population in Canada and the United States. The Delphi technique in this validation study was valuable since it facilitated the examination, by a large group of nurses with varied education and experience, of what constitutes quality nursing care across a variety of settings.

Content validation of the criteria and standards through the use of the Delphi provided Manitoba nurses with a base from which to proceed to the next step in the evaluation of quality nursing care. This next step, or second task, is the quantification of the described indicators of the construct. To this end, instruments are now being developed* and their validity and reliability are being tested across a variety of settings.

* Scherer, K., & Cameron, C. A project to measure the quality of nursing practice in the province of Manitoba is in progress and is funded by Health and Welfare Canada (NHRDP), Project No. 6607-1238-46.

REFERENCE NOTE

1. Shields, M. E. Personal communication, June 3, 1982.

REFERENCES

- Alberta Association of Registered Nurses. *Nursing practice standards*. Edmonton, 1981.
- Anastasi, A. *Psychological testing* (3rd ed.). New York: MacMillan, 1968.
- Bender, D. A., Strack, A. E., Ebright, G. W. & von Haunalter, G. Delphi study examines developments in medicine. *Futures*, June 1969, 289-303.
- Bloch, D. Criteria, standards, nurses — crucial terms in quality assurance. *Journal of Nursing Administration*, 1977, 7(7), 20-30.
- Bramwell, L., & Hykawy, E. The Delphi technique: A possible tool for predicting future events in nursing education. *Nursing Papers*, 1974, 6(1), 23-32.
- Canadian Nurses Association. *A definition of nursing practice; Standards for nursing practice*. 1980.
- Chagnon, M. Méthode d'appréciation de la qualité des soins infirmiers. *Systems, Sciences and Health Care*. (edited by Tilquin, Charles), Toronto: Permagon Press, 1977.
- College of Nurses of Ontario. *Standards of nursing practice for registered nurses and registered nursing assistants*. Toronto, 1976.
- Donabedian, A. Evaluating the quality of medical care. *Milbank Memorial Fund Quarterly*, 1966, 44(3), 166-206.
- Hastings, G.E., Sonneborn, R., Lee, G. H., Vick, L., & Sasmor, L. Peer review checklist: Reproducibility and validity of a method for evaluating the quality of ambulatory care. *American Journal of Public Health*, 1980, 70(3), 222-228.
- Hegyvary, S. T. Nursing Process: The basis for evaluating the quality of nursing care. *International Nursing Review*, 1979, 26(4), 113-116.
- Holmes, C. L., Kane, R. L., Ford, M., & Fowler, J. Toward the measurement of primary care. *Milbank Memorial Fund Quarterly*, 1978, 56(2), 231-252.
- Inui, T. S., Hill, T. A., & Leiby, G. M. Hospital screening for tuberculosis: A quality assurance trial. *Medical Care*, 1979, 17(4), 347-354.
- Lemieux-Charles, L. *Blueprint for Nursing*. Toronto: College of Nurses of Ontario, 1980.
- Lindeman, C. A. Delphi survey of priorities in clinical nursing research. *Nursing Research*, 1975, 24(6), 434-441.
- Linstone, H. A., & Turoff, M. *The Delphi method: Techniques and applications*. London: Addison-Wesley, 1975.
- Manitoba Association of Registered Nurses. *Standards of nursing practice*. 1977.
- Manitoba Association of Registered Nurses. *Standards of nursing care*. 1981.
- Milholland, A. V., Wheeler, S. G., & Heieck, J. J. Medical assessment by a Delphi group opinion technic. *New England Journal of Medicine*, 1973, 288(24), 1272-1275.
- Oberst, M. Priorities in cancer nursing research. *Cancer Nursing*, 1978, 1(4), 281-290.
- Osborne, C. E., & Thompson, H. C. Criteria for evaluation of ambulatory child health care by chart audit: Development and testing of a methodology. Final report of the joint committee on quality assurance of ambulatory health care for children and youth. *Pediatrics*, 1975, 56(4), 625-692.

- Romm, F. J., & Hulka, B. S. Developing criteria for quality of care assessment: Effect of the Delphi technique. *Health Services Research*, 1979, 14(4), 309-312.
- Scherer, K., Cameron, C., Ramsay, J., Vogt, C., & Farrell, P. *Methodological considerations in the construction of national data bases: Identification of and access to Canadian nurse educators and researchers*. Proceedings of the International Research Conference, Edinburgh, September, 1981.
- Scherer, K., Cameron, C., & Farrell, P. *Summary report: Responses to the Delphi survey to validate the Manitoba Association of Registered Nurses Standards of Nursing Care*. Proceedings of the Nursing Research Conference, Victoria, June 1982.
- Ventura, M. R. & Waligora-Serafin, B. Study priorities identified by nurses in mental health settings. *International Journal of Nursing Studies*, 1981, 18, 41-46.
- Weaver, T. The Delphi method: Some theoretical considerations. In Ziegler and Marnen (Eds.). *The potential of educational futures*. Worthington, Ohio: Charles B. Slack, 1972.
- Williamson, J. W. *Assessing and improving health care outcomes*. Cambridge, Massachusetts: Ballinger, 1978.

RÉSUMÉ

La technique Delphi comme méthode de sélection des critères d'évaluation des soins infirmiers

La technique Delphi est une méthode de rassemblement et de globalisation systématique de jugements raisonnés d'un groupe de spécialistes sur des questions et des problèmes particuliers. Son élaboration et son emploi dans les domaines commercial, éducatif, militaire et médical ont été bien documentés au cours des trois dernières décennies. La technique Delphi ne semble pas encore avoir été utilisée dans la sélection et la validation des critères d'évaluation des soins infirmiers. Le présent article décrit l'utilisation de la technique Delphi dans l'élaboration et la validation des normes génériques des soins infirmiers au Manitoba. L'opinion selon laquelle tous les infirmiers qui exercent leur profession sont des spécialistes constitue une façon originale d'appliquer la technique Delphi à ce projet. De plus, la stratification de ce groupe d'experts dans la réalisation de l'échantillonnage a amélioré l'analyse Delphi. Enfin, la gradation des questions a réduit au minimum les erreurs systématiques communes aux enquêtes.