Policy, Politics, & Nursing Practice

Building Consensus Using the Policy Delphi Method

Mary Kay Rayens and Ellen J. Hahn

The online version of this article can be found at:

Published by:

$SAGE

http://www.sagepublications.com

Additional services and information for *Policy, Politics, & Nursing Practice* can be found at:

Email Alerts: http://ppn.sagepub.com/cgi/alerts

Subscriptions: http://ppn.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://ppn.sagepub.com/content/1/4/308.refs.html

Building Consensus Using the Policy Delphi Method

Mary Kay Rayens, PhD
Ellen J. Hahn, DNS, RN

This article describes the use of the policy Delphi method in building consensus for public policy and proposes a technique for measuring the degree of consensus. The policy Delphi method is a systematic method for obtaining, exchanging, and developing informed opinion on an issue. It can be used to develop consensus either for or against policy issues. The method includes a multistage process involving the initial measurement of opinions (first stage), followed by data analysis, design of a new questionnaire, and a second measurement of opinions (second stage). The interquartile deviation is presented as one way of measuring consensus, and the McNemar test is described as a way to quantify the degree of shift in responses from the first to second stage. The application of the method is illustrated by a case example from a study of state legislators' views on tobacco policy.

Building consensus is an essential component of any policy-making process. The hallmarks of the policy Delphi method are to bring together stakeholders with opposing views and to systematically attempt to facilitate consensus as well as to identify divergence of opinion (Strauss & Zeigler, 1975). As many health policy issues are complex, the policy Delphi method is an appropriate tool because it can address a multiplicity of issues and provide direction for policy changes (Critcher & Gladstone, 1998). Unfortunately, this method has not been widely used or reported in the literature (Critcher & Gladstone, 1998; Linstone & Turoff, 1975).

The purposes of this article are to describe the use of the policy Delphi method in building consensus for public policy and to propose a technique for measuring the degree of consensus. The application of the method is illustrated by a case example from a study of state legislators' views on tobacco policy (Hahn, Toumey, Rayens, & McCoy, 1999). Because tobacco control policy development is highly contentious, particularly in tobacco-growing states, the policy Delphi method is well suited for building consensus on tobacco policy issues.

The policy Delphi method is a systematic, intuitive forecasting procedure used to obtain, exchange, and develop informed opinion on a particular topic. Intuitive forecasting procedures are best suited for complex problems for which policy alternatives are not well defined and for which theories or empirical data are not available to make a forecast (Dunn, 1994). This method was developed by the RAND Corporation in the late 1940s (Dalkey & Helmer, 1963) and has been used to study consensus on a variety of issues including drug policy (Rainhorn, Brudon-Jakobowicz, & Reich, 1994), educational issues (Cookson, 1986; Raskin, 1994), nursing administration management issues (Jairath & Weinstein, 1994), and military policies (Linstone & Turoff, 1975).

Characteristics of the Policy Delphi Method

The goals of the policy Delphi method are to describe a variety of alternatives to a policy issue (Strauss & Zeigler, 1975) and to provide a constructive forum in which consensus may occur. The policy Delphi method is a multistage process involving the initial measurement of opinions (first stage), followed by data analysis, design of a new questionnaire based on group response to the previous questions, and a second measurement of opinions (second stage; McKenna, 1994). Statistical group feedback—information about the beliefs of other participants during the first-stage interview—is used in the second-stage interview to facilitate consensus on policy beliefs. Panels of experts or key stakeholders are participants in developing the content of the questionnaire and in responding to issue items. This process allows participants to reconsider their opinions in light of the views of other stakeholders and can be repeated until consensus is reached or saturation of opinion occurs. The number of stages may range between two and five (Critcher & Gladstone, 1998). The policy Delphi method's unique strength is that it incorporates education and consensus-building into the multistage process of data collection, thus enabling description of agreement about specific policy options among key players in the policy decision process. Taking part in the Delphi process can be a highly motivating experience for participants.

Although most applications of the policy Delphi method rely on written questionnaires, some use in-person individual or group interviews, phone or e-mail interviews, or computer conferencing procedures (Dunn, 1994). In-person interviews greatly increase participation (McKenna, 1989) and investment in the project. The use of face-to-face interviewing is especially appropriate with participants who are in leadership positions because their time may be very limited.

The participants in the policy Delphi process should be selected to represent a wide range of opinions (Dunn, 1994). Depending on the policy issue area, the number and type of participants will vary. A typical policy Delphi sample size may range from 10 to 30 participants (Dunn, 1994). As the complexity of the policy issue increases, the sample size needs to be larger to include the entire range of participants both for and against the policy issue area. The type of participants selected includes both formal and informal stakeholders who have vested interest in the policy issue. These participants have varying degrees of influence, hold a variety of positions, and are affiliated with different groups.

First-stage policy Delphi questions typically include four categories of items: forecast, issue, goal, and options (Dunn, 1994). Forecast items provide the participant with a statistic or estimate of a future event. Participants are asked to judge the reliability of the information presented. For issue items, respondents rank issues in terms of their importance relative to others. Goal items elicit opinions about the desirability of certain policy goals. For options items, respondents identify the likelihood that specific options might be feasible policy goals. Because policy Delphi questions are designed to elicit conflict and disagreement as well as to clarify opinions, the response categories do not typically permit neutral answers. The response choices are often rated on a 4-point Likert-type scale. The response choices for forecast items range from *certainly reliable* to *unreliable*. For issue items, response categories range from *very important* to *unimportant*. The response choices for goal items range from *very desirable* to *very undesirable*. For option items, the range is from *definitely feasible* to *definitely unfeasible*.

Depending on the policy area and the level of expertise of the researcher conducting the policy Delphi study, the specific items are developed by the participants or the researcher or a combination

TABLE 1:   Examples of Delphi Items and the Distribution of Responses for First and Second Stages

Item	Response	n	Percentage	IQD
First stage				
Reliability: "More tobacco jobs have been lost due to imports of foreign tobacco or overseas manufacturing than from antismoking efforts. In your opinion, how reliable is this statement?"	Very reliable	13	11.3	1.00
	Reliable	64	55.7	
	Unreliable	38	33.0	
Desirability: "One goal of state policy might be to provide technical and financial support for tobacco farmers who are interested in farm diversification. How desirable is this objective?"	Very desirable	57	49.1	1.00
	Desirable	45	38.8	
	Undesirable	13	11.2	
	Very undesirable	1	0.9	
Feasibility: "The state could provide low-interest loans to assist farmers to supplement their income with other crops. How likely is it that the General Assembly would pass such a policy?"	Very likely	10	8.6	1.00
	Likely	47	40.5	
	Unlikely	57	49.1	
	Very unlikely	2	1.7	
Second stage				
Feasibility: "In this study, 88% of the current members agreed that Kentucky should support farm diversification; 49% thought it is likely the state would pass a law to offer low-interest loans for farm diversification. How likely is it that the General Assembly would pass such a policy?"	Very likely	24	20.9	0.00
	Likely	70	60.9	
	Unlikely	18	15.7	
	Very unlikely	3	2.6	

NOTE: IQD = Interquartile deviation.

of both (Dunn, 1994). If the researcher is not aware of the full range of policy issues, the first-stage interview may be more open ended, in which the participants identify and rank the relevant policy issues. On the other hand, if the researcher is familiar with the policy area under consideration, the first-stage interview items may be entirely specified by the researcher. However, the participants always have the opportunity to add or delete policy issue areas during the first-stage interview process.

Following the first stage, the data are analyzed to determine participants' positions on each interview item. Measures of both central tendency and variability are used to summarize opinions. Based on these measurements, some items are omitted from subsequent stages due to lack of variability in response. In other words, items for which consensus has been achieved are not included in subsequent stages. Items for which there is a lack of agreement among participants are included in subsequent stages. Criteria for determining con-

sensus are discussed below. Summary information for each item from the previous stage is used to frame the subsequent interview item. Participants are then asked to reconsider the desirability of a specific goal in light of the views of the group as a whole. Table 1 displays an example of first- and second-stage items and demonstrates how data collected at the first stage are incorporated into the second-stage interview guide.

When the policy Delphi process is complete, participants are informed of the convergence and divergence of opinions that have occurred during the course of the study.

Measuring Degree of Consensus and Assessing Shift in Opinion

The approach to measuring consensus is the least-developed component of the policy Delphi method (Crisp, Pelletier, Duffield, Adams, & Nagy, 1997), and it varies from study to study. Frequency distributions are often used to assess agreement (McKenna, 1994), and the criterion of

at least 51% responding to any given response category is used to determine consensus (McKenna, 1989). In one study using yes-no response categories, the criterion for agreement was 67% of participants giving the same response (Alexandrov, Pullicino, Meslin, & Norris, 1996). Many researchers ask respondents to rank or weight ideas or issues. Mean rankings and standard deviations are calculated, with a decrease in standard deviation between stages indicating an increase in agreement (Hakim & Weinblatt, 1993; Jairath & Weinstein, 1994). Some investigators ask panels of experts to prioritize ideas by assigning a rank score, but they analyze the responses using qualitative methods (Cookson, 1986; Jairath & Weinstein, 1994). Others use interquartile deviation (IQD) to determine consensus. This method was chosen for use in the case example reported in this article. The interquartile range is the absolute value of the difference between the 75th and 25th percentiles, with smaller values indicating higher degrees of consensus. Raskin (1994) identified an IQD of 1.00 or less as an indicator of consensus. Spinelli (1983) considered a change of more than 1 IQD point in each successive stage as the criterion for convergence of opinion. Clearly, there is no consensus in the literature about how to use or interpret IQD as a method of data analysis for the policy Delphi process. The potential range of IQD values depends on the number of response choices, with larger IQDs expected as the number of response choices increases. Thus, the use of a particular IQD as a cutoff for consensus requires consideration of the number of response choices. In this article, a strategy for use and interpretation of the IQD is proposed.

In most studies using the policy Delphi method, shift in opinion from first to second stage is assessed using qualitative methods. As an alternative, the McNemar test may be used to quantify the degree of shift in responses from the first to second stage (Hahn et al., 1999). This test is a modification of the paired *t* test and was developed for use with categorical data (McNemar, 1947). The McNemar test, which is from the family of chi-square tests, determines whether the percentage of respondents who become more positive on a given item differs significantly from the percentage who become more negative. The application of the McNemar test is described in the following case example.

CASE EXAMPLE

The policy Delphi method was used in a study of Kentucky state legislators' views on tobacco policy (Hahn et al., 1999). The purpose of the study was to describe the level of consensus among Kentucky legislators with regard to tobacco control and tobacco farming policy and to assess the degree of shift toward concurrence on tobacco policy. Kentucky leads the nation in both tobacco use and burley tobacco production. Almost one third of Kentucky adults (30.8%) smoke, compared to less than one fourth of adults in the United States (23.2%; Centers for Disease Control and Prevention, 1998). Tobacco is the state's primary cash crop, yielding more than $730 million in 1997 (U.S. Department of Agriculture, 1999). Similar to other tobacco-growing states, Kentucky has few tobacco control laws, and they are relatively weak compared to those in non-tobacco-growing states (Moore, Wolfe, Lindes, & Douglas, 1994; Welch, 1999). For example, the average cigarette tax in the six major tobacco-growing states is $.07 per pack, whereas the national average is $.419 per pack (Welch, 1999). Given the complex and controversial nature of tobacco policy in Kentucky, the policy Delphi method is an appropriate tool for determining and facilitating consensus among policy makers.

The policy Delphi process for this study was limited to two stages due to potential for participant attrition as well as budget constraints. In-person interviews were used to maximize participation. Persistent phone recruitment was helpful in scheduling times that were convenient to the legislator. All 138 members of the 1998 Kentucky General Assembly were invited to participate due to the complexity of the policy issue area and the need for a larger sample size to gauge a wide range of views both for and against the policy alternatives. Prior to the 1998 Kentucky General Assembly, 116 lawmakers (84.1%) participated in first-stage policy Delphi interviews, and all but one of them took part in the second stage. Participants and nonparticipants did not differ on party affiliation or house membership (House vs.

Senate). The first- and second-stage interviews were completed by the same experienced male interviewer in the legislators' offices. The first-stage interviews lasted approximately 30 minutes, and the second-stage lasted approximately 20 minutes. Each interview item was read aloud, and the lawmaker was asked to respond. Given that the participants in this study were very busy people in leadership positions, it was important to use a procedure that facilitated participation.

Interview Guide

A 57-item interview guide was developed based on current issue areas in tobacco control policy and developments in the tobacco-farming situation. Key stakeholders reviewed an initial draft of the interview guide for appropriate content and policy alternatives. In addition, the interview guide was revised based on pilot testing with 30 former Kentucky legislators (Hahn & Rayens, 1999). Due to the strong correlation between responses to issue and goal items in the pilot study, the interview guide was modified to include only three categories of items: forecast, goal, and option. This reduced the number of first-stage items from 84 in the pilot study to 57 in this case example, diminishing the potential for response burden.

The category names of forecast, goal, and option were changed to reliability, desirability, and feasibility items, respectively, to reflect the nature of the response choices and to create labels that were more intuitive. Legislators rated the reliability of information related to the tobacco policy options using a 4-point Likert-type scale ranging from *certainly reliable* to *very unreliable*. Next, their views on the desirability and feasibility of these policy options were obtained using a similar scale ranging from either *very desirable* to *very undesirable* or *very likely* to *very unlikely* (see Table 1).

The second-stage interview guide included the desirability and feasibility items for which there was not agreement during the first stage. Because they were not linked with specific policy alternatives, reliability items were not included in the second stage, which streamlined this interview. Additional policy option items were included in the second-stage interview based on suggestions from participants during the first-stage interviews and new issues that emerged after the first stage.

For example, a desirability item was added to reflect the suggestion of one legislator that the Medicaid formulary be expanded to cover the costs of voluntary smoking cessation programs. Several items pertaining to the tobacco settlement between state attorneys general and the tobacco companies also were added in the second stage.

Measuring Consensus

Items with IQD = 0.00 were considered to reflect consensus and were not included in the second-stage interview guide. Some items with IQD = 1.00 also were omitted in the second stage because there was a high degree of agreement among respondents. Other items with IQD = 1.00 were included in the second-stage interview guide because there was considerable variability in the distribution of responses among these items. For example, as illustrated in Table 1, 88% were generally positive (i.e., *desirable* or *very desirable*) about the state's providing technical and financial support for tobacco farmers interested in farm diversification. Only 49% thought it likely or very likely that the state would pass a law to offer low-interest loans for farm diversification. Although these two items demonstrated markedly different degrees of consensus, they both had IQDs of 1.00 (see Figure 1). On the desirability item, there was a difference of 76% between the percentage of respondents who were generally positive and the percentage who were generally negative toward providing technical and financial support for farm diversification, reflecting a high degree of consensus. In contrast, there was a difference of only 2% between the percentage of respondents who were generally negative and the percentage who were generally positive about the likelihood that the state would provide low-interest loans for farm diversification.

Because the IQD method lacked sensitivity in distinguishing degree of agreement for items with IQD = 1.00, a secondary criterion for determining consensus for these items was developed. Items with IQD = 1.00 for which the percentage of generally positive respondents was between 40 and 60 were determined to indicate lack of agreement and were retained for the second-stage interview. Thus, items with IQDs of 1.00 and with more than 60% of respondents answering either generally positive or generally negative were considered to
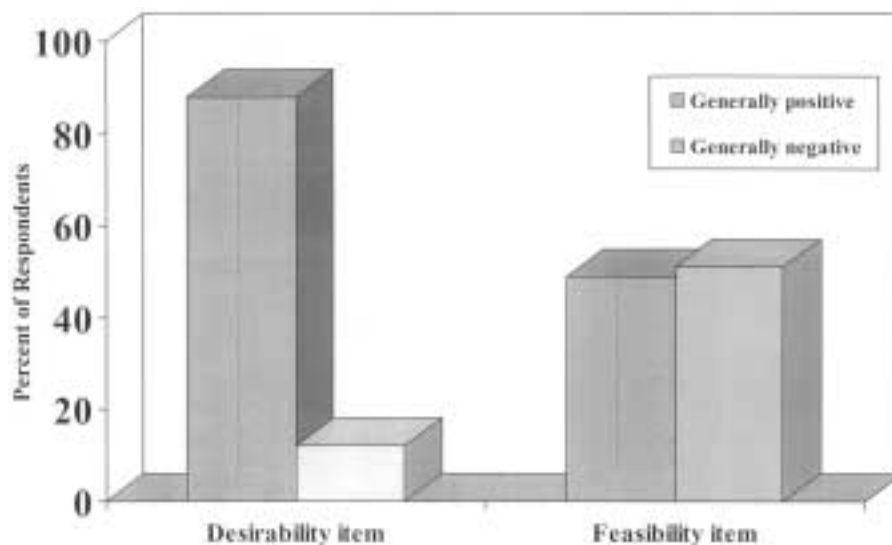
Figure 1:   Two Items With IQD = 1.00: Desirability and Feasibility for Farm Diversification Policies, First Stage (*N* = 116)
NOTE: IQD = interquartile deviation.

be in agreement and were not included on the second-stage interview. This analysis strategy is somewhat similar to that developed by Alexandrov et al. (1996), who used a cutoff of 67% in one of two categories (e.g., yes-no) to designate consensus. In this case example, a slightly less stringent cutoff was chosen to minimize the response burden for the participants. This selection method allowed the testing of the policy Delphi method with the subset of first-stage items on which the legislators had the least agreement. Only items for which there was a lack of agreement (IQD = 1.00 to 3.00, and a maximum percentage positive or percentage negative of less than 60% with IQD = 1.00) were included on the second-stage interview. The use of both criteria is necessary because it is possible, for example, that the frequency distribution of responses to a given item might be bimodal, resulting in an IQD greater than 1.00, although 60% of the respondents are generally positive toward the policy issue. In this case, the item would be included on the second-stage interview because consensus was not attained during the first stage.

A total of 40 desirability and feasibility items were included on the first-stage interview. Of these, 29 demonstrated consensus using the dual criteria previously outlined and were not included on the second-stage interview. Of the 11 items for which consensus was not reached in the first

stage, legislators were in consensus on two items after the second-stage interview.

Degree of Shift From First to Second Stage

The McNemar test was used to determine degree of shift from first to second stage. Of the 11 interview items common to both stages, 5 demonstrated a significant shift from the first to the second stage. On 3 of the 5 items, the legislators became more positive from the first to the second stage; on 2 of the items, lawmakers became more negative. Legislators' views did not shift significantly from first to second stage for the remaining 6 items common to both interviews. For example, 50% of lawmakers became more positive and only 10% became more negative toward the feasibility of providing low-interest loans for farm diversification (Feasibility Item 1; see Figure 2). This degree of shift yielded a significant McNemar test ($\chi^2 = 23.1$, $p < .001$). The remaining 40% of legislators did not change their views on the feasibility of providing low-interest loans for farm diversification. Table 2 displays the pattern of responses in each of the four response categories at the two stages for this item. As an example of an item without significant shift from first to second stage, 25% of lawmakers became more positive and 22% became more negative toward the likelihood that the state would require tobacco companies to
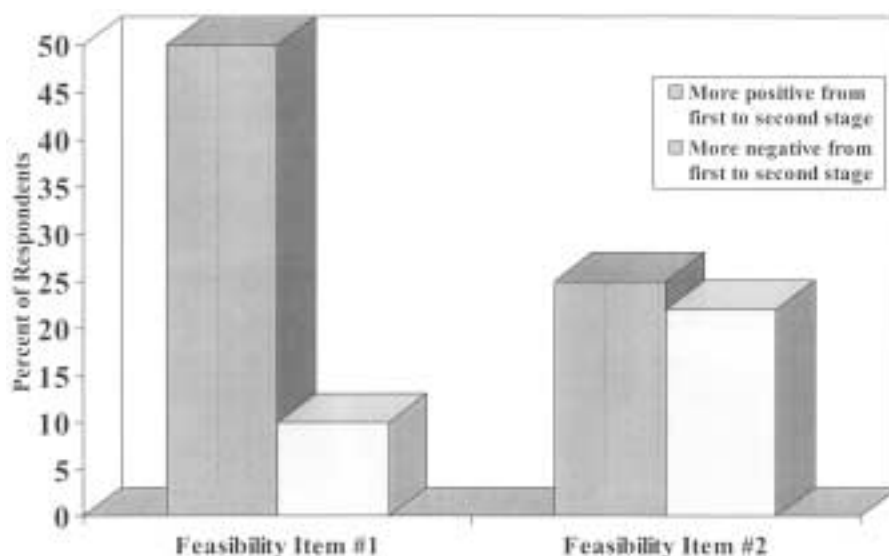
Figure 2:   **Shift of Opinions From First to Second Stage on Two Feasibility Items (*n* = 115)**

contribute funds for farm diversification (Feasibility Item 2; χ^2= 1.3, p > .10). The remaining 53% of lawmakers did not change their views on this issue.

The measure of consensus and the assessment of shift are two distinct concepts. It is possible to have consensus at the second stage without having a significant shift in opinion and vice versa. For example, of the five items that demonstrated significant shift from first to second stage, consensus at second stage was attained for only one. Conversely, there was a second stage item that reached consensus for which the McNemar test of shift was not significant.

SUMMARY

The policy Delphi method is a useful tool for systematically building consensus among decision makers, especially when policy alternatives are not well defined and the issues are complex. The policy Delphi method facilitates the development of consensus either for or against policy issues and should not be confused with lobbying. Although there are a variety of policy Delphi modalities (e.g., phone, written surveys) used to interact with participants, face-to-face interviewing may enhance the involvement of and partici-

TABLE 2:   **Relationship Between First- and Second-Stage Responses for Feasibility Item 1**

	Second Stage			
	Very Likely	Likely	Unlikely	Very Unlikely
First stage				
Very likely	5	2	3	0
Likely	11	31	4	1
Unlikely	8	36	10	2
Very unlikely	0	1	1	0

NOTE: Feasibility Item 1 demonstrated a significant shift. The number of respondents with a given response pattern is recorded in each cell.

pation by elected officials. The study described in the case example used a modified set of interview item categories due to the high degree of correlation between goal and option items found in an earlier pilot study with a similar population using a similar interview guide.

We recommend the use of the IQD approach to data analysis in policy Delphi studies as an objective and rigorous way of determining consensus. Although IQDs have been used in other studies to assess consensus, we propose that an IQD of 1.00 may be an insufficient criterion for determination of agreement, especially with only four response

categories per item. For those items with an IQD of 1.00, we suggest examining the proportion of responses that are generally positive and using a predetermined cutoff (less than 40% or more than 60% in this study) to ascertain consensus. The McNemar test for assessing shift is sensitive to changes of opinion in either a positive or a negative direction. We recommend the use of the McNemar test for assessing shift of opinion because it is consistent with the goal of the policy Delphi method: to arrive at group consensus on a variety of policy issues rather than to push policy in one direction or the other.

REFERENCES

Alexandrov, A. V., Pullicino, P. M., Meslin, E. M., & Norris, J. W. (1996). Agreement on disease-specific criteria for do-not-resuscitate orders in acute stroke. *Stroke*, *27*, 232-237.

Centers for Disease Control and Prevention. (1998). State-specific prevalence among adults of current cigarette smoking and smokeless tobacco use and per capita tax-paid sales of cigarettes—United States, 1997. *Morbidity and Mortality Weekly Report*, *47*, 922-926.

Cookson, P. S. (1986). Charting the unknown: Delphi and policy Delphi strategies for international co-operation. *International Journal of Lifelong Education*, *5*(1), 3-13.

Crisp, J., Pelletier, D., Duffield, C., Adams, A., & Nagy, S. (1997). The Delphi method? *Nursing Research*, *46*, 116-118.

Critcher, C., & Gladstone, B. (1998). Utilizing the Delphi technique in policy discussion: A case study of a privatized utility in Britain. *Public Administration*, *76*, 431-449.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, *9*, 458-467.

Dunn, W. N. (1994). *Public policy analysis: An introduction* (2nd ed.). Englewood Cliffs, NJ: PrenticeHall.

Hahn, E. J., & Rayens, M. K. (1999). Consensus for tobacco policy among former state legislators using the policy Delphi method. *Tobacco Control*, *8*, 137-140.

Hahn, E. J., Toumey, C. P., Rayens, M. K., & McCoy, C. (1999). Kentucky legislators' views on tobacco. *American Journal of Preventive Medicine*, *16*(2), 81-88.

Hakim, S., & Weinblatt, J. (1993). The Delphi process as a tool for decision making. *Evaluation and Program Planning*, *16*, 25-38.

Jairath, N., & Weinstein, J. (1994). The Delphi methodology: Part 1. A useful administrative approach. *Canadian Journal of Nursing Administration*, *7*(3), 29-42.

Linstone, H. A., & Turoff, M. (Eds.). (1975). *The Delphi method: Techniques and application*. Reading, MA: Addison-Wesley.

McKenna, H. P. (1989). The selection by ward managers of an appropriate nursing model for long-stay psychiatric patient care. *Journal of Advanced Nursing*, *14*, 762-775.

McKenna, H. P. (1994). The Delphi technique: A worthwhile research approach for nursing? *Journal of Advanced Nursing*, *19*, 1221-1225.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*, 153-157.

Moore, S., Wolfe, S. M., Lindes, D., & Douglas, C. E. (1994). Epidemiology of failed tobacco control legislation. *Journal of the American Medical Association*, *272*, 1171-1175.

Rainhorn, J. D., Brudon-Jakobowicz, P., & Reich, M. R. (1994). Priorities for pharmaceutical policies in developing countries: Results of a Delphi survey. *Bulletin of the World Health Organization*, *72*, 257-264.

Raskin, M. S. (1994). The Delphi study in field instruction revisited: Expert consensus on issues and research priorities. *Journal of Social Work Education*, *30*, 75-89.

Spinelli, T. (1983). The Delphi decision-making process. *Journal of Psychology*, *113*, 73-80.

Strauss, H. J., & Zeigler, L. H. (1975). The Delphi technique and its uses in social science research. *Journal of Creative Behavior*, *9*(4), 253-259.

U.S. Department of Agriculture. (1999). *Economic Research Service* [Online]. Available: http://www.econ.ag.gov/briefing/tobacco.

Welch, C. E. (Ed.) (1999). *State legislated actions on tobacco issues*. Washington, DC: American Lung Association.

Mary Kay Rayens, PhD, is a research assistant professor in the Colleges of Nursing and Medicine at the University of Kentucky and is the associate director of the Biostatistics Consulting Unit at the University of Kentucky Chandler Medical Center. She received a PhD in statistics from the University of Kentucky in 1993. Her research interests include the design and analysis of longitudinal studies for both clinical trials and surveys. With Dr. Ellen J. Hahn, she leads the program evaluation team for the Kentucky Tobacco Cessation and Prevention Program at the Kentucky Department for Public Health. Their research includes policy surveillance and monitoring of tobacco policies in schools, worksites, health facilities, and food service establishments.

Ellen J. Hahn, DNS, RN, is an associate professor in the College of Nursing at the University of Kentucky. She also is an associate in the University of Kentucky Center for Health Services Management and Research. She received her DNS in health policy and health of the community from Indiana University in 1992. Dr. Hahn is involved in alcohol, tobacco, and other drug prevention policy research. With Dr. Rayens, she leads the program evaluation team for the Kentucky Tobacco Cessation and Prevention Program at the Kentucky Department for Public Health. Their research includes policy surveillance and monitoring of tobacco policies in schools, worksites, health facilities, and food service establishments.