IN MEMORIAM

Florence Downs.

EDITORIAL

Judith Vessey, PhD, RN, FAAN, Associate Editor; Susan Gennaro, DSN, RN, FAAN, Associate Editor

FEATURES

The Staff Nurse Decisional Involvement Scale: Report of Psychometric Assessments.
Donna Sullivan Havens; Joseph Vasey

The Magnet Process and the Perceived Work Environment of Nurses.
Jeannie P. Cimiotti; Patricia M. Quinlan; Elaine L. Larson; Diane K. Pastor; Susan X. Lin; Patricia W. Stone

Effects of Three Groin Compression Methods on Patient Discomfort, Distress, and Vascular Complications Following a Percutaneous Coronary Intervention Procedure.
Linda L. Chlan; Julie Sabo; Kay Savik

Application of the CuSum Technique to Evaluate Changes in Recruitment Strategies.
Patrick McNees; Karen Hassey Dow; Victoria Wochna Loerzel

METHODS

Multilevel Modeling of a Clustered Continuous Outcome: Nurses’ Work Hours and Burnout.
Sunhee Park; Eileen T. Lake

Evaluating Content Validity for Children’s Self-Report Instruments Using Children as Content Experts.
Janet L. Stewart; Mary R. Lynn; Merle H. Mishel

Erratum

Veronica F. Rempusheski; Catherine T. O'Hara

Indexes

PDF Only

Indexes.
Florence Downs—a well-recognized nursing leader, educator, editor, and scholar—helped shape nursing as an intellectual discipline and wrote extensively about the importance of links between research and practice.

To read more about Florence Downs and her contributions to nursing, see Fairman and Mahon. (2001) Oral history of Florence Downs. Nurs Res. 50, 322–328.
When we were contemplating writing this editorial, memories and ideas sprouted freely, intertwined with each other like tendrils of wild vines. But taming these thoughts into a coherent eulogy was not easy without the editor.

In recognition of her many scholarly contributions, this tribute is to Florence Downs. She is well known as the first academic editor of *Nursing Research*, a position she held from 1979 to 1997. Less appreciated is her private side. She was a talented cook, an expert seamstress, a professionally trained coloratura soprano, and closest to her heart, a master gardener. Providing the rich loam needed by young seedlings so they could flourish and produce luxurious blooms was highly rewarding for her. Florence was always more interested in nurturing perennials rather than annuals, whether they were plants or people. She invested her energies in helping researchers become skilled and prolific writers so they might advance a body of knowledge, rather than taking the easier path of editing selected manuscripts merely to ensure that they were suitable for publication.

An author of over nine books, 50 articles, and 80 editorials, Florence delighted in a well-turned phrase and carefully executed text. But she had little patience with muddied prose. Manuscripts seen as in lamentable condition were subjected to the same “tough love” as a weedy patch in her garden. Misplaced phrases, dangling participles, and the like were attacked like underground runners of tenacious weeds, eliminated so they could not sneakily invade an idea, choking out the meaning of the text. Although many felt the pain of having their words pruned—that sometimes felt more akin to being “trimmed” by a backhoe—inevitably the result was a more robust cohesive message, rooted in theory and logic to withstand the winds of criticism.

For those of us who worked with Florence, our indelible memory is of a woman sitting at her desk surrounded by a jar of sharpened pencils and a pile of manuscripts awaiting attention. What was less known was the path that Florence took to become a dedicated nurse scholar. Florence began her nursing career in 1944 when she graduated from St. Luke’s Hospital. Like many nurses of her generation, her mother was a nurse and Florence entered nursing during World War II, so that she could participate in the nurse cadet corps. Recognizing her strength was in nurturing the human condition, Florence chose psychiatric nursing as her specialty and she used her communication skills throughout her entire career.

In the 1970s, a bumper sticker proclaimed: “Grow where you are planted.” Florence took this message to heart. As a native of the metropolitan New York City area, she began her academic career cultivating nursing knowledge while she served as the Director of Postgraduate and Research Programs at New York University. Florence was instrumental in advancing the University’s Doctor of Philosophy in Nursing program and received the first federal funding for doctoral education in nursing. Later, she transplanted a hybrid of this successful program to the University of Pennsylvania. She then went on to advance nursing research and doctoral education across the globe. Florence chaired over 100 dissertations; many of these graduates have gone on to conduct original programs of research, edit nursing journals, and lead the profession through academic, clinical, and administrative roles.

Florence saw that the future for nursing lay in the nurturing of its science. When Florence assumed the editorship of *Nursing Research*, the readership was only 6,600, the journal was in financial difficulty, and few original research studies were published. With few existent doctoral
programs, few established programs of nursing research, and few external organizations to fund nursing research, Florence’s gifts, including her unwavering belief in the importance of nursing science being used at the bedside, were well timed to help move nursing science forward. Today, there are more than 80 nursing doctoral programs, greater than 34 million dollars in federal funding dedicated to nursing research, and over 20 nursing research journals, with virtually all clinical journals publishing research or evidence-based practice articles. It is hard to think that as little as three decades ago, nursing research was merely a seed of unrecognized potential.

After her “retirement,” Florence melded all of her talents into a final landscape befitting a lifetime of commitment and service. Not yet ready just to reap the rewards of her long career, Florence planted her last perennials. She continued to cultivate the skills needed for scholarly inquiry with doctoral students while simultaneously planting a contemplative garden outside the door of her church’s rectory. She continued to edit scholarly manuscripts, but also served as a part-time parish receptionist, where she edited the church bulletin much to the pastor’s bemusement. Last fall she was inducted as a “Living Legend” into the American Academy of Nursing. Her final legacy is endowing the Florence and William Downs Professorship in Nursing dedicated to advancing nursing knowledge through research at New York University, College of Nursing.

Whether it was plants or words, Florence’s greatest joy was envisioning how the mysteries they contained would be revealed in time. She had the patience to plan for the next season. She was never cowed by a challenge and was willing to plant even in the rockiest of ground. She was a woman of faith who never questioned why the most beautiful roses often had the fiercest thorns. She had a terrific sense of humor. She never stopped fighting with the weeds and insects that are always ready to attack, but she consistently laughed at the unique and different challenges they offered to her expertise. We will all continue to enjoy her many gardens, but the gardener will be sadly missed.

Judith Vessey, PhD, RN, FAAN
Susan Gennaro, DSN, RN, FAAN
Associate Editors
Nursing Research, 1988 to 1997
The Staff Nurse Decisional Involvement Scale

Report of Psychometric Assessments

Donna Sullivan Havens • Joseph Vasey

Background: For decades, enhancing staff nurse decisional involvement in matters of nursing practice and patient care has been identified as a long-term strategy to improve the quality of the nursing work environment and the safety and quality of patient care.

Objective: To describe psychometric assessments of the Decisional Involvement Scale (DIS), a diagnostic and evaluative measure of nurse decisional involvement.

Methods: A series of assessments were conducted to evaluate the psychometric performance of the scale. Content validity was assessed by experts in the field. Descriptive statistics were used to examine the use and performance of the scale. The contrasted groups approach was used to assess construct validity. Item analysis was used to explore evidence of the internal consistency of items and subscales across multiple samples. Structural modeling was used to conduct a confirmatory factor analysis using data from two independent samples of staff registered nurses (RNs; n = 849 and 650).

Results: Acceptable content validity indexes (CVIs) were independently generated by three content experts. Construct validity was supported, as hypothesized; nurses working on professional practice units scored significantly higher for all items when compared to nurses working on units without professional practice models in place. Internal consistency (coefficient alpha) was high and nearly identical for the total measure and all subscales across the two independent nurse samples. Six subscales were identified using factor analysis, and these were confirmed by structural modeling.

Conclusion: Psychometric findings support that the DIS is a valid and reliable measure of staff nurse decisional involvement.

Key Words: decision making • scale development • shared governance

Once again in the early 2000s, the nation is confronting dissatisfaction of many nurses, a critical nurse workforce shortage, and calls for enhanced quality of patient care. In addition, a growing body of research presents convincing evidence that the way nurses are organized affects the quality of the working environment and nurse, patient, and organizational outcomes. For instance, features of professional nursing practice models such as a high level of nurse decisional involvement have been empirically associated with better outcomes.

Thus, it is not surprising that major organizations and legislators are urging implementation of organizational models that enhance staff nurse decisional involvement. For instance, the Nurse Reinvestment Act (2002) proposes incentives for hospitals to “improve the retention of nurses and enhance patient care . . . by [among other strategies] promoting nurse involvement in the organizational and clinical decision making processes of the healthcare facility.” The American Nurses Credentialing Center’s Magnet Recognition ProgramTM promotes nursing involvement in decisions about nursing practice and hospital policy. Implementation of organizational features such as decisional involvement that make magnet hospitals successful has been encouraged by the American Hospital Association (2002), the American Nurses Association (2002), the American Association of Critical-Care Nurses (2005), and the Joint Commission on the Accreditation of Healthcare Organizations (2002). In a recent report by the Institute of Medicine (2003), increased nurse involvement in decision making was identified as a major factor in enhancing patient safety. These calls echo what organizational experts (Begun, 1985; Heydebrand, 1983; Scott, 1982), government officials (Kusserow, 1988), and nurse leaders (Aydelotte, 1981, 1983; Maas & Jacox, 1977; McClure, Poulin, Sovie, & Wandelt, 1983; Prescott & Dennis, 1985) have advocated for decades, the need to organize nursing practice in hospitals to enhance staff nurse influence on practice and hospital policy.

Central to any initiative to enhance nurse involvement in decision making is the ability to measure staff nurse decisional involvement. The Decisional Involvement Scale...
Decisional Involvement

Decisional involvement is defined as the pattern of distribution of authority for decisions and activities that govern nursing practice policy and the practice environment. Development of the DIS was grounded in two bodies of literature on organizationally based professionals: sociology of the professions and professional nursing practice models. Specifically, development of the DIS was grounded in the literature that emphasizes the organization of professional work and professional control over work and working conditions. The work of Scott (1982) and Aydelotte (1981, 1983) is classic in terms of describing arrangements for organizing professionals who work in organizations and professional nursing practice, and served as the guiding framework for the development of the DIS when it was developed in 1990.

From a sociologic perspective, Scott (1982) described three models for structuring the work of professionals who work in healthcare organizations: (a) the autonomous model (administrators delegate the control of most of the professional activities within the organization to the professional staff); (b) the heteronomous model (administrators retain control over most professional activities with elaborate sets of rules and regulations and routine supervision; for example, professional employees are clearly subordinated to an administrative framework with minimal autonomy); and (c) the conjoint model (professional participants and administrators are equal in the power they command and in the importance of their spheres of action). According to Scott, the heteronomous model was the typical model for structuring the work of nurses in hospitals. In contrast, in the conjoint model, nurses and administrators coexist in a state of collaboration, interdependence, and mutual influence; professionals and administrators each serve as the dominant force in certain areas. For example, professionals maintain responsibility for the care of clients and administrators provide the resources to shape the optimal environment needed by professionals to meet client goals. According to Scott, this arrangement promotes increased collaboration between administrators and practitioners to develop and meet organizational goals and it recognizes the autonomy of professionals and the interdependence that professionals and administrators share, as well as the interdependence that nursing shares with other healthcare disciplines.

Aydelotte (1981, 1983) proposed a model for nursing practice similar to Scott’s conjoint model, identifying three domains for policy development and administration in nursing: (a) professional nurses have sole authority and responsibility (matters related to patient care and its improvement, certifications, and performance standards for practitioners); (b) administration has sole authority and responsibility (resource acquisition and allocation, interdepartmental, and institutional relations); and (c) the two share authority and responsibility (policy development and administration of needed resources, scheduling, cost saving, support services, general personnel policies, and the work environment). Aydelotte referred to this arrangement as a professional organizational model that was needed for professional nursing practice. These models of collaborative relationships between administrators and professionals provided the framework for the development of the DIS items. The overriding perspective used to shape the development of the DIS was that of Pointer (1976), which encourages managing with professionals versus managing of professionals.

The Decisional Involvement Scale

The DIS is a tool that can be used in multiple ways to plan and evaluate change in the organization of nursing to (a) measure perceived actual levels of decisional involvement, (b) assess desired levels of decisional involvement, (c) measure decisional dissonance (a gap between actual and desired degree of decisional involvement), (d) identify concordance between staff and management perceptions regarding actual and preferred levels of decisional involvement, (e) identify areas for change, and (f) monitor the effects of strategies implemented to enhance staff nurse decisional involvement. Using the DIS in a program of benchmarking, feedback, and open dialogue between staff and management or administration is suggested as a strategy for implementation of professional practice.

The DIS consists of 21 items that measure actual and desired decisional involvement of staff registered nurses (RNs) on a nursing unit related to six constructs (subscales) empirically identified through factor analysis: unit staffing, quality of professional practice, professional recruitment, unit governance and leadership, quality of support staff practice, and collaboration or liaison activities. The DIS uses a 5-point scale to indicate the degree to which decisions are the responsibility of staff nurses and administration/management (actual decisional involvement). For each item, respondents indicate which nursing group (staff nurses or administration/management) they perceive as having the primary responsibility for the decision or activity on their unit. Response choices are as follows: administration/management only = 1; primarily administration/management with some staff nurse input = 2; equally shared by administration/management and staff nurses = 3; primarily staff nurses with some administration/management input = 4; and staff nurses only = 5. Results can be considered by individual items, by subscale, or as an overall scale. High scores suggest a high degree of staff RN involvement, a low score suggests a low degree of staff RN involvement, and a midrange score suggests sharing of decision making between administration or management and staff RNs. Another form elicits responses to indicate which group staff nurses perceive should have the primary responsibility for the decision or activity (desired decisional involvement)
involvement). Decisional dissonance, the discordance between actual and desired levels of decisional involvement, is assessed by calculating the absolute difference between actual and desired scores. See Havens and Vasey (2003) for more detailed information about the DIS.

**Psychometric Assessment of the DIS**

Unless otherwise specified, analyses were conducted using data reflecting staff nurse reports of desired decisional involvement. The Desired-Practice Scale was selected for analysis because it represents coherent attitudes about how nurses perceive that decision making should be managed. In contrast, when responding to items about actual practices, nurses might be reporting on behaviors that, for a variety of institution-specific, idiosyncratic reasons, might not reflect a consistent representation of the underlying construct.

**Validity**

Validity refers to the “determination of whether or not a device or method ... measures what it purports to measure” (Waltz, Strickland, & Lenz, 1991, p. 3).

**Content Validity**

Content validity refers to the content representativeness or relevance of the items in a measure (Lynn, 1986). Content validity of the DIS was established during scale development by following a two-stage development and judgment process (Lynn, 1986; Waltz et al., 1991). First, items were generated following a comprehensive review of the literature on the organization of professionals working in organizations and professional nursing practice models. During this phase, the content domain was identified, the items were generated, and the first draft of the instrument was formulated.

Content validity was further established by a second phase: judgment quantification by content experts that the instrument was content valid. In this phase, three nurse content experts, all doctoral-prepared nurse researchers who had extensively studied and published on staff nurse decisional involvement in the hospital setting, were presented with a specific set of instructions to use when assessing the content relevance of the items and the scale as a whole. Each one of the experts independently reviewed and rated each item on content relevance, judging the items on quality and representativeness of the domain and the objective for which they were written. They also provided input regarding omitted items they viewed as important.

On the second iteration of the review, each expert completed a review of the items and the instrument and conducted a content validity assessment; each expert independently produced a content validity index (CVI) of 1.0 (Havens, 1990). Because the CVI represents the proportion of items that are evaluated as relevant and strongly representing their construct, the findings provide robust support that the items are representative measures of their respective constructs (Waltz et al., 1991).

**Construct Validity**

Construct validity is an instrument’s ability to “… measure the attribute of interest” (Waltz et al., 1991, p. 143), in this case the degree of staff nurse decisional involvement. To assess the construct validity of the DIS, a confirmatory factor analysis of the instrument was evaluated for two independent samples of staff RNs (n = 849 and 650) working on a variety of acute care units in two academic health centers in two states (Massachusetts and North Carolina). The DIS was administered through internal department of nursing assessments to identify areas for improvement. Hospital researchers shared with the authors data collected from all RN staff completing the surveys, although no respondent demographic characteristics were available. SPSS Amos version 5.0 was used to conduct latent construct structural modeling. Confirmatory factor analysis permits an examination of the simultaneous performance of all six subscales and their corresponding items. To the extent that the model can replicate the relationships in the actual data, construct validity is demonstrated. The structural model is based on the hypothesized relationships between items and constructs. The observed response on each item has two components—the construct that underlies the item and measurement error.

Two issues are addressed using findings regarding the structural model. First, does the model adequately replicate the relationships in the data? This is a question of the degree to which the relationships specified by the theoretical model (between items and subscales and between the subscales themselves) match observed relationships in the data. Goodness of fit can be assessed in multiple and often competing ways (Bentler & Bonett, 1980; Browne & Cudek, 1993). Atheoretic measures of fit, including discrepancy measures ($c^2_{min}/df$ or $c^2_{max}/df$) and the root mean square error of approximation (RMSEA) test the hypothesis that the relationships specified by the model are inconsistent with the data. The $c^2_{max}/df$, also known as relative chi-square, is the minimum discrepancy function divided by the degrees of freedom. The ratio should be close to 1.0 for correct models, but there is considerable disagreement of what an acceptable value is (Carmines & McIver, 1981). Marsh and Hocevar (1985) summarize recommendations from various researchers regarding the

### Table 1. Selected Measures of Fit for Specified, Saturated, and Independence Measurement Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Atheoretic Measures of Fit</th>
<th>Comparison to Baseline Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T^2$ $c^2_{min}/df$</td>
<td>RMSEA</td>
</tr>
<tr>
<td>Sample A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Six-scale</td>
<td>5.160</td>
<td>0.070</td>
</tr>
<tr>
<td>Independent</td>
<td>181.807</td>
<td>0.528</td>
</tr>
<tr>
<td>Sample B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Six-scale</td>
<td>4.266</td>
<td>0.071</td>
</tr>
<tr>
<td>Independent</td>
<td>181.807</td>
<td>0.528</td>
</tr>
</tbody>
</table>

Note. 1998 data. Scales as defined by Havens (1990). NFI indicates normal fit index; RMSEA, root mean square error of approximation.
relative chi-square measure ($c_{\text{min}}/df$) that indicate that, while values in the range of 2–3 represent good fit, values as high as 5.0 may be reasonable. Wheaton, Muthen, Alwin, and Summers (1977) suggested a ratio of 5.0 or less as "reasonable." Carmines and McIver (1981) argue that values less than 3.0 are acceptable but Byrne (1989) believes that a ratio greater than 2.0 is inadequate.

The normed fit index (NFI) can serve as an indicator that estimates where the measurement model falls between a poorly fitting independence model (values close to 0) and a perfectly fitting saturated model (values close to 1). The NFI compares the minimum discrepancy function of the model under consideration to the minimum discrepancy function of the independence model. As such, it can be interpreted as a ratio that indicates how far the specified model is between the poorly fitting independence model and the perfectly fitting saturated model. The closer NFI is to 1.0, the nearer the specified model is to the saturated model. Bentler and Bonett (1980) state that models with indices less than .90 can be substantially improved. A theoretic and normed fit measures of fit for the DIS measurement model are shown in Table 1.

The measures reported in Table 1 show that the six-scale model approximates the performance of a saturated model or perfectly explanatory model rather than an independence model. The chi-square value ($c_{\text{min}}/df$) for both samples is somewhat higher than the rule of thumb suggested by some authors. RMSEA values are approximately .07 for both samples, a value within the range of acceptable fit proposed by Brown and Cudek (1993). NFI values for both samples exceed .98, further suggesting that the models are adequate. These analyses provide evidence of construct validity in that the DIS measurement model closely resembles a saturated model and is clearly superior to an independence model.
The second issue addressed by the structural model is the relationship between subscales and their constituent items. Items should have relatively strong relationships to their subscales and relatively small error terms. This is demonstrated by large factor loadings. Table 2 shows that loadings range from approximately .60 to .95 and are consistent across both samples. These figures are consistent with low measurement error and factor invariance—the items are strongly related to their scales and their performance is reliably repeatable across samples. These findings mirror the results from the single-scale item analyses. Note that factor loadings for the items in Scale 6 are somewhat lower compared to other scales. Again, this is reflected in the alpha coefficients for this subscale (Table 3), which were around .70, substantially below that of the other subscales.

Table 4 presents correlations between the scores on the six latent constructs. These are somewhat higher than the correlations between the observed subscale scores presented in Table 5. They show the relationships between the predicted, error-free measures of the six constructs. While they are slightly higher than the observed score correlations, the pattern of correlations is similar. Further, the correlations obtained from the two nurse samples are similar. This is consistent with a reliable, construct-valid instrument.

Further evidence of construct validity was reported by Somers (1995). The DIS was administered to contrasted groups known through means other than the instrument to have varying degrees of decisional involvement. The DIS actual and desired forms were completed by 131 staff RNs working on 18 medical–surgical nursing units in two large
TABLE 6. Descriptive Statistics for the Six DIS Subscales (Desired)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Statistic</th>
<th>Sample 1</th>
<th>Sample 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decisions related to unit staffing</td>
<td>n</td>
<td>845</td>
<td>646</td>
</tr>
<tr>
<td>Mean</td>
<td>3.22</td>
<td>3.29</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>3.00</td>
<td>3.50</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0.75</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>Decisions related to the quality of professional practice</td>
<td>n</td>
<td>845</td>
<td>646</td>
</tr>
<tr>
<td>Mean</td>
<td>2.98</td>
<td>3.00</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>3.00</td>
<td>3.00</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0.69</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>Decisions related to professional recruitment</td>
<td>n</td>
<td>845</td>
<td>645</td>
</tr>
<tr>
<td>Mean</td>
<td>2.54</td>
<td>2.60</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>2.67</td>
<td>2.67</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>.84</td>
<td>.84</td>
<td></td>
</tr>
<tr>
<td>Decisions related to unit governance and leadership</td>
<td>n</td>
<td>846</td>
<td>645</td>
</tr>
<tr>
<td>Mean</td>
<td>2.72</td>
<td>2.70</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>2.67</td>
<td>2.67</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>.66</td>
<td>.68</td>
<td></td>
</tr>
<tr>
<td>Decisions related to the quality of support staff practice</td>
<td>n</td>
<td>833</td>
<td>636</td>
</tr>
<tr>
<td>Mean</td>
<td>3.01</td>
<td>2.95</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>3.00</td>
<td>3.00</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>.70</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td>Collaboration/liaison activities</td>
<td>n</td>
<td>844</td>
<td>644</td>
</tr>
<tr>
<td>Mean</td>
<td>3.33</td>
<td>3.31</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>3.33</td>
<td>3.33</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>.67</td>
<td>.65</td>
<td></td>
</tr>
</tbody>
</table>

Ideally, the performance of the DIS would be compared with the performance of another scale that is conceptually very similar. However, there is no measure that measures the same constructs at the unit level. Preliminary findings are reported from a study in progress at six rural PA hospitals to provide preliminary evidence of convergent validity. Nurses (n = 1,071) completed the DIS and the Practice Environment Scale (PES; Lake, 2002). The PES includes one subscale that assesses nurse participation in hospital affairs. An examination of the items constituting this subscale suggests that although it does not measure exactly the same constructs as the DIS, it is concerned with a related issue. In preliminary analyses, the DIS actual scales were found to be positively correlated with the PES participation in hospital affairs subscale. Correlations ranged from .21 to .28, all significant at the .001 level. Because higher DIS actual scores mean that nurses are more involved in decision making, and higher PES scores mean that there is greater participation of nurses in hospital operations, the positive correlations are expected, although they are not large.

The DIS desired scales were found to be negatively correlated with the PES participation in hospital affairs subscale. This suggests that lower or poorer participation of nurses in hospital operations (as measured by the PES) is reflected by a desire for more nurse involvement in decision making (as measured by the DIS scales). Correlations ranged from -.11 to -.24, significant at the .001 level. Again, while the correlations are not large, they are in the expected direction.

Reliability

DIS total scale and subscale reliability was assessed in terms of internal consistency (alpha). Analyses were based on data from the two independent RN samples described earlier. Table 6 presents descriptive statistics for the whole scale and each DIS subscale for the two samples. For all subscales, and in both samples, scores tended to tightly cluster around the scale midpoint. For each subscale, average and median (50th percentile) scores are similar. The two samples were consistent in the patterns of score distributions, evidence that the subscales operated in the same manner in both samples. This finding supports the hypothesis that the subscales operate in a consistent manner across samples. High scores suggest that respondents believe there should be a high degree of staff RN involvement in decision making and low scores indicate a preference for administration or management taking responsibility for decision making. Midrange scores, as are strongly evident in both samples, suggest a preference for shared decision making between staff and administration/management.

Table 3 presents the estimates of the assessments of internal consistency for the whole scale and the six DIS subscales for the two RN samples. As seen in Table 3, coefficient alpha values are consistently high for the total scale and all subscales across both samples. Within scales, item-total correlations were similar across items. The results show that each subscale demonstrates coherence; constituent items work well together to provide a good estimate of the construct that they claim to measure. Further,
the performance of the subscales is nearly identical across the two independent samples of nurses, implying that internal consistency is a real and repeatable phenomenon.

In an item analysis, each subscale and the items that compose it are considered in isolation from other subscales and items in a instrument. Table 3 shows that the DIS subscales are moderately correlated with one another and that the pattern of correlations is the same across both samples. The correlations presented in Table 3 also show that while the subscales covary, or overlap with one another, to some extent, they represent relatively independent measures of the six dimensions.

Discussion

There are limitations in this analysis. For instance, in several cases, the psychometric assessments presented in this article would have benefited from a description of the demographics of the nurse samples. However, the data used in these assessments were provided by hospitals that had administered the instrument to all staff RNs to assess opportunities for change without collecting information about the demographic characteristics of the RNs. Therefore, all of the assessments could not be put in context of the sample, which limits generalizability. However, new research by the study authors will provide data to overcome this limitation.

In this analysis, the focus was on the psychometric assessment of the desired DIS and not the actual version of the scale. This approach was taken because of the belief that the desired scale represents coherent attitudes about how nurses perceive that decision making should be managed. In contrast, it was felt that when responding to items about actual practices, nurses would be reporting on behaviors that, for a variety of institution-specific idiosyncratic reasons, might not reflect a consistent representation of the underlying construct. It may be critically important to understand the notion of concordance between nurse reports of actual versus desired decisional involvement in order to shape work environments.

In the psychometric assessments reported in this article, evidence has been found to support the reliability and validity of the DIS. Subjective ratings by a panel of subject matter experts indicated that all items are good exemplars of their associated constructs. Additionally, agreement between the experts was high. Analysis of the instrument’s performance in nursing environments provided further support. First, mean response profiles were similar across two independent samples of nurses working under similar conditions. Item analyses showed high levels of internal consistency for all scales and for the instrument as a whole. Finally, confirmatory factor analyses demonstrated that the theoretical factorial structure of the instrument is very successful at replicating the observed relationships between items in two independent samples of respondents. Taken together, these findings present promising evidence that the DIS is a reliable and valid instrument for measuring nurses’ attitudes and beliefs about their involvement in work-unit decision making.

The literature presents evidence that organizing nursing practice to increase staff nurse involvement in decisions about the content and content of practice produces positive outcomes for patients, staff, and organizations. While there are other measures that may be used to assess related concepts such as shared governance, nurse autonomy, and nurse control over practice, the DIS is unique because it addresses staff nurse actual and desired involvement in specific decisions and it was developed to be applicable and easily understood by those at the forefront of shaping the practice environment at the unit level.

References


The Magnet Process and the Perceived Work Environment of Nurses

Jeannie P. Cimiotti • Patricia M. Quinlan • Elaine L. Larson • Diane K. Pastor
Susan X. Lin • Patricia W. Stone

Background: Several studies have described the work environment of nurses from magnet and nonmagnet hospitals, but there have been no studies of nurses from hospitals in the magnet application process.

Objectives: To compare the differences between characteristics of hospitals and nurses from three hospital types: magnet hospitals, hospitals in the process of applying for magnet certification, and nonmagnet hospitals, and how nurses from these hospitals perceive their work environment.

Methods: In a national, cross-sectional survey of critical care nurses, the Perceived Nursing Work Environment (PNWE) instrument was administered to measure nurses’ perceptions of their work environment.

Results: Data were available from 2,092 nurse surveys. Over a third of the respondents were from in-process hospitals and almost half were from nonmagnet hospitals. The majority of nurses were female and from large hospitals in the Atlantic region. The mean age of nurses was 39.5 years and the mean years of work experience in the intensive care unit (ICU) was 10.2 years. Higher nurse scores were significantly associated with magnet certification on one subscale of the PNWE, nursing competence.

Discussion: Nurses from magnet hospitals had a positive perception of nursing competence in their work environment. Further research is necessary to examine the nurse work environment and to determine if the characteristics of magnet hospitals have changed.

Key Words: magnet hospitals • nurses • work environment

A magnet hospital has organizational, leadership, and professional practice characteristics that are consistent with a discrete set of research-based attributes associated with positive nursing and patient outcomes, more so than other hospitals. The concept of a magnet hospital was developed during the 1980s when a task force appointed by the American Nurses Association (ANA) observed that some hospitals outperformed others in their ability to recruit and retain nurses (McClure, Poulin, Sovie, & Wandelt, 1983). This landmark study identified and described characteristics that now represent the cornerstone of credentialing standards used for determining magnet status. Among the characteristics identified, autonomy and control over practice were noted as critical to sustaining professional vitality. McClure and Hinshaw (2002) identified distinct qualities described by nurses that made working at certain hospitals more desirable.

Recent studies have examined the relationship between the perceived work environment of nurses at magnet and nonmagnet hospitals (Kramer & Schmalenberg, 2003a, 2003b; Laschinger, Almost, & Tuer-Hodes, 2003; Upenieks, 2002, 2003). No studies were found that examine nurse perceptions at hospitals in the magnet certification process or to determine changes that occur when hospitals acquire the characteristics of magnetism. This comparative study examined the differences between characteristics of hospitals and how these nurses perceive their work environment in three hospital types: magnet hospitals, hospitals in the magnet certification process, and nonmagnet hospitals.

The Magnet Recognition Program was established in 1993. The American Nurses Credentialing Center (ANCC) is the oversight division of the ANA and evaluates hospitals seeking magnet certification. The Standards for Nurse Administrators (ANCC, 2003) is the foundational document upon which the magnet program is based. This document outlines the credentialing standards that hospitals must meet to achieve magnet designation. The standards are based upon the organizational elements otherwise known as the forces of magnetism identified in the original magnet study. Initially, the basic characteristics were categorized into three major areas: administration, quality leadership, and organizational structure. Today, the standard measures of excellence are categorized by the Magnet Recognition Program into four major areas based upon the...
findings of the original study and subsequent research of magnet hospital characteristics: (a) management, philosophy, and practice; (b) integration of recognized quality improvement standards; (c) support for professional practice and continued competence; and (d) understanding and respecting cultural and ethnic diversity of patients, significant others, and healthcare providers (ANCC, 2003).

Magnet application and designation is a staged process that consists of (a) application, (b) documentation review, (c) facility site visit, and (d) review and decision. Each phase of the process is rigorous and follows successful completion of the preceding phase. The process takes about 2 years. Magnet designation is granted for a 4-year period, after which hospitals must reapply. Interim statistical reports to ANCC are required annually.

Over the last 2 decades, there has been considerable study of magnet hospitals by those interested in understanding the relationship of nurse and patient outcomes to the workforce environment. Kramer and Hafner (1989) advanced previous research by developing the Nursing Work Index (NWI), designed to quantify the values nurses associate with baseline magnet characteristics. Kramer and Hafner first used the NWI to study the relationships among organizational characteristics, job satisfaction, and productivity. They compared the NWI ratings of magnet hospital nurse participants with their nonmagnet counterparts. Reports of greater job satisfaction by magnet nurses paved the way for further study in this area and the identification of additional magnet characteristics, in particular competence as the baseline for autonomy (Kramer, 1990; Kramer & Schmaling, 1991a, 1991b).

During the mid-1990s, Aiken, Smith, and Lake (1994) worked toward continued understanding of the relationship between outcomes and nursing practice by comparing magnet and nonmagnet hospitals. Lower mortality and greater patient satisfaction were found in magnet hospitals. In this research, magnet attributes such as autonomy, control over practice, and physician–nurse relationships were identified. These attributes were measured by the Nursing Work Index—Revised (NWI-R), which was based upon the foundational subscales of the original NWI (Aiken & Patrician, 2000). Lake (2002) furthered this work by conducting a factor analysis, which resulted in a 48-item measure, the Practice Environment Scale (PES).

Another team of researchers used the NWI-R to further understand critical care nurses’ perception of their work environment (Choi, Bakken, Larson, Du, & Stone, 2004). This study led to the development of the Perceived Nursing Work Environment (PNWE), an updated version of the NWI-R. The PNWE includes an additional subscale, a positive scheduling climate, as an important characteristic of the work environment in hospital settings.

Methods

Hospital Sampling

This study was part of a larger project funded by the Agency for Healthcare Research and Quality that examined registered nurse working conditions and patient outcomes in intensive care units (ICUs) in the United States (P. Stone, Principal Investigator, ICU Working Conditions). Hospitals were invited to participate through the Association for Professionals in Infection Control and Healthcare Epidemiology, Inc. (APIC) listserv and through the Centers for Disease Control and Prevention’s National Nosocomial Infections Surveillance (NNIS) System. The NNIS system is a voluntary surveillance and reporting system that provides infection control practitioners with a set of definitions for healthcare-associated infections, risk factors, and populations monitored (Horan & Emori, 1997). Hospitals were eligible if they collected infection surveillance data using the NNIS definitions in one or more of the following adult ICU types: coronary care, medical, surgical, and medical–surgical.

Hospital characteristics such as bed size, region, and teaching status were obtained from the American Hospital Association database. Enrolled hospitals were then classified into Atlantic, Central, and Pacific geographical regions. Determination of each hospital’s magnet status was made by asking each Chief Nursing Officer if the institution had received magnet certification, was applying for magnet certification, or did not have magnet certification as of December 31, 2002. Institutional Review Board approval was obtained from participating hospitals as well as from the Columbia University Medical Center.

Nurse Sampling

Surveys were distributed using guidelines from previous research (Smith, Salyer, Geddes, & Mark, 1998). Over a 6-month period (October 2002 to April 2003), nurse surveys were distributed in bulk to a site coordinator at each hospital. The nursing leaders of each ICU identified the number of registered nurses working in each enrolled unit before survey distribution. The survey was coded to the ICU and a stamped return envelope was provided to each survey respondent to assure anonymity. After 2 months, if an ICU had less than a 50% response rate, a second round of surveys was mailed to that ICU with a request for each nurse who had not already completed the survey to please do so. To encourage the nurses to complete and return the surveys, separate (to maintain anonymity) postcards were distributed. If the postcard was returned, the respondent was entered into lotteries of gift certificates and small prizes.

Registered Nurse Survey Measure

Nurses were surveyed using the 42-item PNWE and nine demographic questions (Choi et al., 2004). Respondents were asked their perception of their work environment by indicating 1 (strongly agree) to 4 (strongly disagree) on a Likert scale. Scores for negative items were reverse-coded, so that higher scores indicated positive perceptions. Table 1 is a descriptive summary of the subscales of the PNWE. The reliability of the PNWE is evidenced by an overall Cronbach’s alpha of .95. Six of the seven subscales have moderate to high alphas (.70 to .91): professional practice, staffing and resource adequacy, nursing management, nursing process, nurse/physician collaboration and nursing competence. Positive scheduling climate, a construct unique to the PNWE, had a lower alpha (.56) possibly due to the fact that this is a three item subscale.
TABLE 1. A Descriptive Summary of the Seven Subscales of the Perceived Nursing Work Environment Instrument

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Description</th>
<th>α</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional practice</td>
<td>Opportunities for nurse advancement and involvement in hospital governance</td>
<td>.91</td>
<td>13</td>
</tr>
<tr>
<td>Staffing and resource adequacy</td>
<td>Enough nurses, support staff and time to provide quality patient care</td>
<td>.83</td>
<td>5</td>
</tr>
<tr>
<td>Nurse management</td>
<td>Nurse leaders are available and supportive of nursing staff</td>
<td>.88</td>
<td>5</td>
</tr>
<tr>
<td>Nursing process</td>
<td>Nursing diagnoses, care plans and team nursing are used to deliver quality care</td>
<td>.70</td>
<td>6</td>
</tr>
<tr>
<td>Nurse–physician collaboration</td>
<td>Teamwork and collaboration between nurses and physicians to provide quality care</td>
<td>.84</td>
<td>4</td>
</tr>
<tr>
<td>Nurse competence</td>
<td>Orientation and preceptor programs; experienced and clinically competent nurses</td>
<td>.72</td>
<td>6</td>
</tr>
<tr>
<td>Positive scheduling climate</td>
<td>Participation in developing flexible schedules; permanent nurses never float</td>
<td>.56</td>
<td>3</td>
</tr>
</tbody>
</table>

Data Analysis

Chi-square tests were used to compare the proportion of nurse respondents in the three types of hospitals by gender, education, hospital bed size, hospital region, hospital teaching status, and NNIS status. To determine the association between magnet status and the continuous variables (age of nurses, years in the healthcare profession, years in the ICU, job years, and the PNWE subscales), a general linear model (GLM) was used. The GLM method was appropriate due to lack of normal distribution in some variables (Agresti & Finlay, 1997).

Because the nurses in a single ICU are likely to perceive similar work environments and these data are likely to be clustered, seven (one for each PNWE subscale) multivariate, multilevel analyses were conducted also. The multivariate, multilevel analyses allow for control of potential confounding variables (such as years in the ICU) while controlling for clustering. Those variables with a p value < .25 (Hosmer & Lemeshow, 2000) in the univariate models were included in multivariate, multilevel models to determine independent associations with the scores of the seven subscales. In these analyses, data on years worked in the ICU were ranked into quartiles with the 25%, 50%, and 75% cut-points used in analysis. To prevent spurious results due to multicollinearity, a correlation matrix of the continuous variables was examined. Variables with a multiple correlation coefficient > .70 were excluded from the models. All data were analyzed using SAS version 9.0 for Windows (SAS Systems, Inc., Cary, NC).

Results

Data were available on the survey responses of 2,323 nurses from 110 ICUs in 68 hospitals. Data on magnet status were available from 107 ICUs in 64 hospitals: 8 had magnet certification, 21 were in the magnet process, and 35 were nonmagnet hospitals. The final sample size of usable nurse surveys was 2,092, the mean number of nurse respondents per ICU was 33, and the average response rate was 41%. The number of ICUs per hospital ranged 1–5; the majority of hospitals (59%; n = 38) reported one ICU.

Hospital and Nurse Characteristics

A summary of hospital and nurse characteristics of the survey respondents is shown in Table 2. Thirteen percent of the nurses were from magnet hospitals, 37.6% were from hospitals applying for magnet status, and almost half were from nonmagnet hospitals. A total of 878 (42%) nurse respondents were from the Atlantic region, 696 (33%) from the Central region, and 518 (25%) from the Pacific region. Most (55%) respondents were employed in large (>400 beds) hospitals and were from hospitals that participated in NNIS (71%). There were statistically significant differences among the three types of hospitals for region, bed size, teaching status, and NNIS status (all p < .0001).

The majority (90%) of nurses were female and the mean age was 39.5 years (SD = 9.41). On average, nurse respondents had worked 15.6 years (SD = 9.30) in the healthcare profession, 10.3 years (SD = 8.48) specifically in the ICU, and 8 years (SD = 7.50) in their current job. There were no statistically significant differences in nurse characteristics among the three types of hospitals.

Work Environment

Table 3 is a comparison of nurse scores on the PNWE subscales for the three types of hospitals. There were significant differences (p < .05) between the mean scores on six out of seven subscales. There were no statistically significant differences on the scores for nurse–physician collaboration. On average, nurses working in hospitals in the magnet application process had significantly lower mean scores on nurse management, staffing and resource adequacy, professional practice, nursing process, and positive scheduling climate when compared to nurses from magnet and nonmagnet hospitals (all p values < .05). Nurses employed in nonmagnet hospitals had lower mean scores on nursing competence when compared to magnet hospitals and those in the application process. Nurses employed in magnet hospitals had higher mean scores on four subscales: professional practice, nursing process, nursing competence, and positive scheduling climate.

Multivariate, Multilevel Analyses

A summary of the multivariate, multilevel analyses is provided in Table 4. Age, years in healthcare, and years on the job were highly correlated with years worked in the ICU and were therefore excluded from these models (r = .71, .83, and .72, respectively). ICU type and ICU years
were included in the models only if they were significant in univariate analysis. In the multilevel models, after adjusting for variability between the hospitals, magnet status was significantly associated with nursing competence \(p = .02\). There was no independent association between magnet status and the remaining six subscales.

In these multivariate, multilevel analyses, other nursing and hospital characteristics were significantly associated with the perceptions of the work environment. In all models, the random effect of the hospital was significantly related \(p < .0001\) to the perceptions of the work environment. Nurses from medical and medical–surgical ICUs perceived higher staffing and resource adequacy \(p < .01; M = 2.64\) and 2.65, respectively) than nurses from coronary and surgical ICUs \(M = 2.52\) and 2.51, respectively). The number of years the nurse worked in the ICU was significantly associated (all \(p \) values < .01) with the mean nurse score on five of the seven subscales: professional practice, staffing and resource adequacy, nursing management, nursing process, and nursing competence. In a post hoc analysis (data not displayed), mean nurse scores on five practice environment factors (professional practice, staffing and resource adequacy, nursing management, nursing process, and nursing competence) were significantly higher (all \(p \) values < .05) for nurses with \(\geq 3\) years working in the ICU when compared to nurses with \(\leq 3\) years working in

<table>
<thead>
<tr>
<th>TABLE 2. Hospital and Nurse Characteristics of Survey Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital characteristic</td>
</tr>
<tr>
<td>Region</td>
</tr>
<tr>
<td>Atlantic</td>
</tr>
<tr>
<td>Central</td>
</tr>
<tr>
<td>Pacific</td>
</tr>
<tr>
<td>Bed size</td>
</tr>
<tr>
<td>(\leq 400)</td>
</tr>
<tr>
<td>(&gt; 400)</td>
</tr>
<tr>
<td>ICU</td>
</tr>
<tr>
<td>Coronary</td>
</tr>
<tr>
<td>Medical</td>
</tr>
<tr>
<td>Surgical</td>
</tr>
<tr>
<td>Medical–surgical</td>
</tr>
<tr>
<td>Teaching hospital</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>NNIS</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Nurse characteristic</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>(&lt;\text{BSN})</td>
</tr>
<tr>
<td>(\geq\text{BSN})</td>
</tr>
<tr>
<td>Age (SD)</td>
</tr>
<tr>
<td>Years in healthcare (SD)</td>
</tr>
<tr>
<td>Years in ICU (SD)</td>
</tr>
<tr>
<td>Years on job (SD)</td>
</tr>
</tbody>
</table>

Notes: Chi-square tests were conducted for categorical variables; general linear models (GLMs) were conducted for continuous variables; \(^{1}\) sample sizes vary due to missing data.
the ICU. There was no association between level of education and any of the subscales of the PNWE.

**Discussion**

This study compared critical care nurses’ perceptions of the work environment from three types of hospitals: magnet, nonmagnet, and those in the process of applying for magnet certification. Most nurse respondents were from the Atlantic region, which is not surprising because the majority of U.S. nurses work in hospitals in that region (Sprately, Johnson, Sochalski, Fritz, & Spencer, 2001). A large number of nurse respondents from magnet hospitals and those from hospitals in the magnet process were NNIS participants, a finding expected as a result of the enrollment process in the primary study. Large numbers of nonmagnet hospital respondents were from small hospitals that were not academically affiliated. Perhaps smaller hospitals do not have the resources necessary to begin the magnet application process.

Upenieks (2003) reported that most nurses at magnet hospitals held bachelors of science in nursing (BSN) degrees and had less nursing experience (<5 years); whereas nurses from nonmagnet hospitals were less educated and had more nursing experience (>20 years). In this study, however, in all three types of hospitals, most nurses were BSN prepared and had an average over 15 years of work experience in healthcare. Because lower levels of education have been associated with nonresponse on self-report instruments (Helasoja, Prattala, Dregval, Pudule, & Kasmel, 2002; Korkkeila et al., 2001), this finding could represent a response bias. However, it might also reflect the critical care sample of nurses.

Nurses from hospitals with magnet certification had higher mean scores on most of the subscales, but surprisingly only nursing competence was statistically significant in the multivariate, multilevel models. An unmeasured difference between hospitals predicted the mean nurse scores on six of the seven subscales of the PNWE, perhaps providing evidence for Kramer and Schmalenberg’s (2004) findings that the NWI no longer captures the essence of magnetism and only measures hospital organizational traits. Also it is likely that the work environment of critical care nurses is so multifaceted that the effect of magnet status among the nurses in these units is not adequately measured. The finding that ICU type was significantly associated with staffing and resource adequacy might reflect the shortage of critical care nurses. Buerhaus, Staiger, and Auerbach (2000) reported that the aging nursing workforce is problematic for ICUs, which typically attract younger nurses.

In a study of magnet hospitals (10 original and 4 ANCC certified), Buchan (1999) found that magnet hospitals have undergone significant reorganization. In these magnet hospitals, skill mix varied significantly; furthermore, continued professional education, professional autonomy, and participative management style were no longer present. The results in this study are consistent with Buchan’s study; when controlling for clustering and other confounding variables, there was no statistically significant association between magnet status and nurse scores for professional practice, nurse management, and nursing process.

While no differences were found in nursing characteristics between groups, there was an association between the number of years worked in the ICU and perception of the work environment. Fewer years in the ICU were consistently associated with significantly higher perceptions of the work environment in five of the seven subscales. This finding may be useful to nurse managers and administrators, and they may be well advised to pay attention to the tenure of their nurses and to look for ways to help more seasoned nurses keep a positive outlook towards their work environment.

Surprisingly, no association was found between education level and nurses’ perception of the work environment. In a study of New York State nurses, it was reported that masters-prepared nurses were significantly more satisfied with their work environment when compared to those with a BSN or lower degree (Ingersoll, Ölsan, Drew-Cates, DeVinney, & Davies, 2002). Nurses with a BSN degree have scored high for values associated with job satisfaction (McNee-Norris & Crook, 2003); nurses with less than a BSN degree have scored low in personal accomplishment, a characteristic that has been linked to burnout (Demir, Ulusoy, & Ulusoy, 2003). More research is needed investigating the relationship among perceptions of work environment, educational preparation, and outcomes.
<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Mean Square</th>
<th>F value</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional practice</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnet i</td>
<td>2</td>
<td>4.07</td>
<td>2.35</td>
<td>.10</td>
</tr>
<tr>
<td>Hospital j(i)</td>
<td>61</td>
<td>2.93</td>
<td>10.33</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ICU type k</td>
<td>3</td>
<td>0.14</td>
<td>0.51</td>
<td>.68</td>
</tr>
<tr>
<td>Education l</td>
<td>1</td>
<td>0.30</td>
<td>1.06</td>
<td>.30</td>
</tr>
<tr>
<td>ICU years m</td>
<td>3</td>
<td>1.29</td>
<td>4.58</td>
<td>.003</td>
</tr>
<tr>
<td>Error (ijkm)</td>
<td>1862</td>
<td>0.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staffing and resource adequacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnet i</td>
<td>2</td>
<td>1.51</td>
<td>0.99</td>
<td>.38</td>
</tr>
<tr>
<td>Hospital j(i)</td>
<td>61</td>
<td>2.48</td>
<td>6.59</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ICU type k</td>
<td>3</td>
<td>1.35</td>
<td>3.58</td>
<td>.01</td>
</tr>
<tr>
<td>Education l</td>
<td>1</td>
<td>0.81</td>
<td>2.15</td>
<td>.14</td>
</tr>
<tr>
<td>ICU years m</td>
<td>3</td>
<td>2.29</td>
<td>6.10</td>
<td>.0004</td>
</tr>
<tr>
<td>Error (ijkm)</td>
<td>1863</td>
<td>0.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nursing management</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnet i</td>
<td>2</td>
<td>1.87</td>
<td>0.86</td>
<td>.43</td>
</tr>
<tr>
<td>Hospital j(i)</td>
<td>61</td>
<td>3.67</td>
<td>8.43</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Education l</td>
<td>1</td>
<td>0.03</td>
<td>0.06</td>
<td>.81</td>
</tr>
<tr>
<td>ICU years m</td>
<td>3</td>
<td>3.13</td>
<td>7.19</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error (ijm)</td>
<td>1865</td>
<td>0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nursing process</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnet i</td>
<td>2</td>
<td>1.24</td>
<td>1.10</td>
<td>.34</td>
</tr>
<tr>
<td>Hospital j(i)</td>
<td>61</td>
<td>1.85</td>
<td>6.23</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ICU years m</td>
<td>3</td>
<td>1.48</td>
<td>4.99</td>
<td>.002</td>
</tr>
<tr>
<td>Error (ijm)</td>
<td>1862</td>
<td>0.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nurse-physician collaboration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnet i</td>
<td>2</td>
<td>0.33</td>
<td>0.34</td>
<td>.71</td>
</tr>
<tr>
<td>Hospital j(i)</td>
<td>61</td>
<td>1.52</td>
<td>4.68</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Education l</td>
<td>1</td>
<td>0.26</td>
<td>0.81</td>
<td>.37</td>
</tr>
<tr>
<td>Error (ij)</td>
<td>2019</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nursing competence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnet i</td>
<td>2</td>
<td>3.59</td>
<td>3.97</td>
<td>.02</td>
</tr>
<tr>
<td>Hospital j(i)</td>
<td>61</td>
<td>1.42</td>
<td>5.08</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ICU type k</td>
<td>3</td>
<td>0.34</td>
<td>1.21</td>
<td>.31</td>
</tr>
<tr>
<td>Education l</td>
<td>1</td>
<td>0.16</td>
<td>0.56</td>
<td>.45</td>
</tr>
<tr>
<td>ICU years m</td>
<td>3</td>
<td>2.27</td>
<td>8.13</td>
<td>.001</td>
</tr>
<tr>
<td>Error (ijkm)</td>
<td>1862</td>
<td>0.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive scheduling climate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnet i</td>
<td>2</td>
<td>4.10</td>
<td>1.62</td>
<td>.21</td>
</tr>
<tr>
<td>Hospital j(i)</td>
<td>61</td>
<td>4.33</td>
<td>12.86</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ICU type k</td>
<td>3</td>
<td>0.13</td>
<td>0.38</td>
<td>.77</td>
</tr>
<tr>
<td>Education l</td>
<td>1</td>
<td>0.60</td>
<td>1.79</td>
<td>.18</td>
</tr>
<tr>
<td>ICU years m</td>
<td>3</td>
<td>0.55</td>
<td>1.64</td>
<td>.18</td>
</tr>
<tr>
<td>Error (ijkm)</td>
<td>1861</td>
<td>0.43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** ICU = intensive care unit. The subscripts for the models are defined as follows: i = magnet group; j = hospital; (ij) = hospital nested within magnet group; k = ICU type; l = level education; m = years worked in ICU; n = error term for variables included in the model.
This study had several limitations. First, because only ICU nurses were surveyed, the results may not be generalizable to other hospital units. Second, our sample of nurse respondents was not randomized and the survey response rate was 41%, which could introduce a measurement bias. Third, the PNWE is a new instrument that warrants further psychometric testing. Fourth, although this study is the first to distinguish nonmagnet hospitals from hospitals in the process of obtaining magnet certification, we were not able to differentiate among the stages of obtaining magnet certification.

Even with these limitations, this study contributes new information to the literature on magnet hospitals. This study compares nurse perceptions of the work environment at magnet hospitals, nonmagnet hospitals, and hospitals in the magnet application process. The magnet status of hospitals was associated with perceptions of nursing competence; and more years worked in the ICU was associated with lower scores in five subscales of the PNWE. Further longitudinal research to examine the magnet process and nurses' perceptions of the work environment is needed to determine if the characteristics of magnet hospitals have changed since first identified 20 years ago and if the process of obtaining magnet accreditation changes the nurse work environment over time.

Accepted for publication May 6, 2005.
This study was funded by the Agency for Healthcare Research and Quality R01 HS13114. The authors acknowledge all of the registered nurses who participated in the survey, as well as site coordinators and nursing administrators from the participating hospitals.

Corresponding author: Patricia W. Stone, PhD, MPH, RN, Columbia University School of Nursing, 617 West 168th Street, New York, NY 10032 (e-mail: ps2024@columbia.edu).

References
Effects of Three Groin Compression Methods on Patient Discomfort, Distress, and Vascular Complications Following a Percutaneous Coronary Intervention Procedure

Linda L. Chlan • Julie Sabo • Kay Savik

Femoral sheath removal followed by compression of the femoral artery after a percutaneous coronary intervention procedure (PCIP) is a nursing responsibility in many acute and critical care settings. Nurses can choose among three accepted methods to achieve hemostasis of the femoral artery: manual pressure or the commercially available Femostop® or C-clamp compression devices. Vascular complications (VCs) arise from compression of the femoral vasculature and occur at rates from 11% to 65% (Lehmann, Heath-Lange, & Ferris, 1999; Pracyk et al., 1998; Rudisill, Williams, Craig, & Schopp, 1997). Evidence has not clearly demonstrated one method to be superior in reducing VCs. Moreover, the process of sheath removal and femoral artery compression can be uncomfortable and distressful for patients. It is not known if one of the compression methods can reduce these patient symptoms. Thus, the purpose of this study was to determine which of the three compression methods causes the least discomfort, is the least distressful for patients, and results in the fewest VCs.

Over one million PCIPs, which include angioplasty, atherectomy, laser angioplasty, stents, and radiation therapy for in-stent restenosis were performed in 2000 as a primary method for treating coronary heart disease (CHD) (American Heart Association, 2004). To gain access to the coronary arteries and cardiac chambers, a femoral arterial sheath is inserted into the groin in the cardiovascular (CV) lab. Anticoagulation is required during PCIPs to prevent acute thrombotic closure of coronary vessels that can occur during insertion of interventional devices. Bed rest is maintained prior to and after femoral sheath removal for 4–6 hr to promote healing of the arterial puncture site.

Linda L. Chlan, PhD, RN, is Associate Professor, School of Nursing; and Kay Savik, MS, is Biostatistician, School of Nursing, University of Minnesota.

Julie Sabo, MN, RN, CCRN, CCNS, is Cardiovascular Clinical Nurse Specialist, John Nasseff Heart Hospital of United Hospital, St. Paul, Minnesota.

Key Words: complications • coronary angioplasty
(Apple & Lindsay, 2000). Hospital length of stay ranges 1–3 days, with a national average of 3 days (American College of Cardiology, 2003).

Primary nursing roles for the care of patients after PCIP are monitoring, preventing, and treating symptoms (e.g., discomfort and distress) and VCs. Femoral sheath removal and compression of the puncture site to achieve hemostasis of the femoral artery following PCIP is a medically delegated nursing responsibility. Nurses at the study hospital choose among three methods to achieve hemostasis (manual pressure, Femostop®, or C-clamp) based on experience and personal bias, not formal protocol. Risks associated with femoral sheath removal include inadequate hemostasis resulting in VCs such as oozing, ecchymosis, or hematoma, development of pseudo-aneurysm, arteriovenous fistulas, thrombosis, thromboemboli, and retroperitoneal bleed (Apple & Lindsay, 2000; Lehmann et al., 1999; Pracyk et al., 1998, 1999; Rudisill et al., 1997). VCs are responsible for increased morbidity, length of stay, patient distress, and discomfort (Pracyk et al., 1998). VC rates following PCIP have been reported to be 11–65% (Kresowik et al., 1991; Lehmann et al., 1999; Pracyk et al., 1998, 1999; Rudisill et al., 1997; Waksman et al., 1995); approximately 2–3% require surgical intervention (Waksman et al., 1995). The intensity of nursing care increases when VCs occur, due to the numerous management interventions required.

Results from randomized trials examining the three compression methods and the incidence of VCs after femoral sheath removal have been mixed. Pneumatic devices like the Femostop® significantly prolong time to hemostasis, promote more incidence of bleeding, and result in larger hematomas, with comparable hematoma (10–13%) and ecchymosis rates (29%–34%) by method (Lehmann et al., 1999). Others have found comparable rates of VCs by compression method (12%) with equal numbers of hematoma (10–13%) and ecchymosis rates (57–61%), or hematoma (up to 65%), but the incidence of vascular pathology (pseudo-aneurysm and arteriovenous fistula) was significantly lower (63% reduction) when a mechanical clamp was used (Pracyk et al., 1998). A meta-analysis of seven published studies documented decreased hematoma formation with mechanical devices, increased incidence of pseudo-aneurysm with manual pressure and C-clamp, no difference in bleeding, and mixed results for time to hemostasis (Jones & McCutcheon, 2002). The investigators concluded that there is a lack of quality, randomized trials evaluating the three compression methods simultaneously with adequately powered sample sizes.

Femoral sheath removal has been reported to cause minimal to moderate discomfort and distress for patients (Fulton et al., 2000; Puntillo et al., 2001). One study reported manual compression as the least painful and Femostop® as the most painful method (Lehmann et al., 1999).

In summary, reports of VC rates following PCIP range 3–65% with inconclusive findings as to which compression method causes the fewest complications. The wide variability in incidence of VCs is attributed to differences among studies in definitions and assessment for VCs and in protocols for sheath removal and groin compression. Furthermore, few investigations were found that examined discomfort and distress associated with compression of the femoral artery. To begin to address these gaps, the primary aims of this study were to determine if one compression method (manual, Femostop®, or C-clamp) after femoral sheath removal following PCIP causes the fewest VCs and less patient discomfort and distress.

**Methods**

**Design, Sample, and Setting**

A three-group experimental design with repeated measures was used. Power analysis calculations based on data from the literature (Bogart, 1995; Pracyk et al., 1998) resulted in the need for a convenience sample of 306 subjects to detect a moderate effect size with significant differences of 60% on discomfort ratings and 20–40% difference among VCs. Participants were recruited from inpatient units and the Same Day Intervention Unit (SDIU) contained in the heart hospital unit of the main hospital from June 2001 through December 2003. The main hospital is a 532-bed tertiary care center located in the urban Midwest. The heart hospital unit consists of 135 beds within the hospital and is a busy comprehensive cardiac care center. In 2002, 7,351 patients were admitted to the heart hospital unit and 1,095 PCIPs were performed. Approval for the use of human subjects in research was obtained from the institutional review boards at the main hospital and the investigators’ university. Investigators approached potential subjects about participation before any anxiolytic medications were administered prior to entering the CV lab.

Any patient admitted to the SDIU or inpatient unit at the heart hospital unit scheduled for an angiogram and possible PCIP who was alert, 18 years of age or older, and could read and write English was invited to participate. Only subjects who had a PCIP while in the CV lab met the inclusion criteria and were admitted to the designated unit after the procedure were included in the final sample. Any patient undergoing a PCIP who was hemodynamically unstable, receiving mechanical ventilation, received thrombolytic therapy within 24 hr before or during the PCIP (streptokinase, alteplase recombinant, reteplase recombinant, tenecteplase, or other thrombolytics), had a known groin pathology, could not read and write English, admitted to an ICU post-procedure, or had a documented mental incompetence in the medical record (e.g., Alzheimer’s disease) was not approached or included in the final sample.

**Procedure**

A group of 30 registered nurses (RNs) from the main hospital were trained as study nurses. The RNs were experienced in the care of CV patients and had further education in the care of post-PCIP patients. They had an average of 12 years of experience. RNs attended an education session on the protocol, demonstration, and verification of skills in all three methods of groin compression and orientation to data collection and data recording. Only trained RNs cared for enrolled study participants.
The second author provided study updates during the course of the subject recruitment, focusing on reviewing the study protocol, required data collection, and spot checks of all study RNs for correct techniques of femoral sheath removal and groin compression.

Before and after the CV lab, these data were abstracted from the medical record: age, gender, ethnicity, current medications, current blood pressure, height, weight, type of PCIP, any complications during the procedure, length of PCIP, size of femoral sheath, coagulation status, medications administered during the PCIP, and medications administered before and after femoral sheath removal.

### Treatment Conditions

A table of random numbers was used to assign participants to one compression method: Femostop®, C-clamp, or manual pressure. Before femoral sheath removal, the RN opened a sealed envelope to determine which compression method would be used for each participant.

The Femostop® is a compression arch with a transparent pneumatic dome. The arch is placed over the patient’s anterior hip area with a mesh belt. The pneumatic dome is placed over the puncture site with the center of the dome 1 cm superior and medial to the puncture site, which places it over the opening in the artery. The dome is inflated with a hand-held manometer while the sheath is removed. Pressure is increased by inflating the dome until hemostasis is achieved; pressure is applied until bleeding is absent.

The C-clamp consists of a flat metal plate (placed under the patient’s buttocks to stabilize the device) and a C-clamp arm. A disposable translucent pad is attached to the tip of the C-clamp arm. Appropriate placement of the translucent pad is achieved in the same manner as the Femostop®. Pressure is placed at the puncture site by lowering the arm of the C-clamp so the pad is centered at the vascular access site. As the femoral sheath is removed, pressure is increased in order to achieve hemostasis; pressure is applied until bleeding is absent.

Manual pressure consists of the nurse using three fingertips to apply direct pressure at the groin puncture site, in the same manner as the mechanical methods. While applying pressure, the femoral sheath is withdrawn and pressure is increased at the groin site until hemostasis is achieved; pressure is applied until bleeding is absent.

Time to hemostasis for each compression method was measured using a stopwatch, beginning when the sheath was withdrawn from the femoral artery. The RN recorded time to hemostasis for each subject on the data log.

### Outcome Measures

Discomfort was defined as the level of uncomfortableness in the groin area as reported by the subject. Discomfort ratings were obtained via a verbal Numeric Rating Scale (NRS). The RN asked each subject to rate the level of discomfort (0 = none to 10 = most uncomfortable ever) in the groin area immediately before femoral sheath removal, 1 min after compression initially applied, 10 min after compression applied, and 10 min after compression released, similar to other investigations (Puntillo et al., 2001).

NRSs have practical advantages over other paper and pencil instruments to measure discomfort in that they are easier to administer and score and have a high rate of correct responses (Frank-Stromborg & Olsen, 1997). These scales are widely used in acute care settings; patients’ verbally rate sensations (e.g., pain) from 0 to 10. Concurrent validity of the discomfort NRS and a 100-mm visual analog scale has been reported as \( r = .86-.95 \), with test–retest reliability of \( r = .97 \) in a sample of emergency department patients (Singer, Kowalska, & Thode, 2001). Construct validity has been reported elsewhere (Puntillo et al., 2001). Correlation among the four discomfort ratings in this study was \( r = .34-.65 \).

Distress was defined as how bothersome the groin compression procedure was to the subject, similar to other investigations of procedural distress (Puntillo et al., 2001). Distress ratings were obtained at the same time as the discomfort ratings using the same NRS. The RN asked each subject to rate the level of distress from 0 = not bothersome at all to 10 = the most bothersome.

The NRS for distress has been used in previous research studies (Puntillo et al., 2001). Procedural distress and pain intensity are related but different experiences for patients and have been found to be correlated \( .51-.67 \) across several medical procedures (Puntillo et al., 2001). Correlation among the four distress ratings in this study was \( r = .30-.65 \). Correlation between discomfort and distress across all time points was \( r = .54 \), ranging \( .26-.63 \).

The RN assessed the groin area for the presence or absence of any VCs prior to sheath removal, immediately after compression released, and 12 and 24 hr after femoral sheath removal. These time points were chosen to coincide with previous work in the area of assessing VCs post-PCIP (Puntillo et al., 2001). VCs included oozing (presence of any leakage of blood from the puncture site), ecchymosis (presence of any skin discoloration without a mass), hematoma (presence of a nonpulsatile mass \( >4 \) cm), or pulsatile mass (presence of a palpable mass with movement corresponding to systole and diastole).

### Data Analysis

Data were entered into a computerized database (Microsoft Access) and then imported to SPSS 11.0 for Windows or SAS V8.0, depending on the desired analysis. Baseline data and demographic variables were assessed via descriptive statistics and exploratory data analysis techniques to discern the shape of the distributions and to assess for any outliers or data entry errors. Time in the CV lab, time to hemostasis, baseline discomfort, and baseline distress were found to be non-normal, necessitating the use of nonparametric statistics in subsequent bivariate analyses. These comparisons between groups were accomplished using a chi-square test of association or a Kruskal–Wallis ANOVA, depending on the level of measurement.

Analysis of the repeated, non-normally distributed continuous variable outcomes (discomfort, distress) and repeated, binary outcomes (presence/absence of VCs) was accomplished using generalized estimating equations (GEE) within the GENMOD procedure in SAS. Distribution of the discomfort and distress ratings required recoding to absent/present to avoid problems with empty cells in the
analysis. Analysis of correlated data that is normally distributed has long been accomplished using generalized linear models. The parameters that result from ordinary or weighted least squares do not apply when longitudinal data that are not normally distributed are considered. Other analyses appropriate for binary data do not account for correlated measures over time, resulting in incorrect parameter estimates due to underestimated standard errors. Hypothesis testing results very often are not replicable. The GEE provides a non-likelihood-based approach to modeling correlated data (Zeger & Liang, 1986, 1992; Zeger, Liang, & Albert, 1988). It also can handle missing data and a mix of binary, categorical, and continuous explanatory variables. Assumptions underlying GEE are that the probability distribution of the data belongs to one of the exponential families (e.g., binomial, Poisson, normal), that a link function can be specified (e.g., logit for binary, identity for normal), and that the covariance structure of the repeated measures can be specified (Zeger & Liang, 1986, 1992; Zeger et al., 1988). An iterative process then ensues until the estimates for beta converge.

For this specific data, the distribution for distress, discomfort, and VCs were not normally distributed. The distribution identified as binary, link function was specified as a logit function, and the correlation between outcome measures was unstructured. With the inclusion of the repeated statement in the GENMOD procedure, responses from different subjects are assumed to be statistically independent and responses from the same subjects are assumed to be correlated (repeats). The independent measure was the type of compression method; the manual method was used as the reference. Other covariates included the measures at baseline for presence of VCs, distress, discomfort, time to hemostasis, and several medications known to influence the outcome measures (epitabiate, heparin, clopidogrel, midazolam, lidocaine, hydromorphone, morphine sulfate, acetaminophen/proxopyhene napsylate, and acetaminophen/ codeine), which were dichotomized yes/no if a subject received the specific medication. Since time to hemostasis (dichotomized a priori as median split) was found to be significantly different between the three compression methods, two GEE models were run for each outcome. The first model included all covariates of interest along with compression method and time to hemostasis as individual variables. Another model was explored with time to hemostasis (dichotomized to <30 and ≥30 min) and compression methods included as their interaction terms only. If none of the interaction terms were significant, the first model was reported.

Results

Demographic and Baseline Data
Mean age of participants was 62.3 (SD = 11.4) with a range of 33–90 years. A majority of participants were male (77%). Participants were Caucasian (n = 295; 96.4%), Hispanic (n = 9; 2.9%), or other (n = 2; 0.7%). Percutaneous transluminal coronary angioplasty (PTCA) with stent placement was the most common PCIP (62%), followed by brachytherapy with stent and PTCA (12%), PTCA alone (11%), stent placement alone (8%), and atherectomy and other (7%). Overall, 56% of participants previously had a PCIP; there was no difference by compression method [\( \chi^2(2) = .51; p = .78 \)]. Participants had a 6.0 FR (n = 275; 90%), 8.0 FR (n = 25; 8%), 4.0 FR (n = 5; 1.7%), or 7 FR (n = 1; 0.3%) femoral sheath placed in the CV lab. There were no documented medical complications for any participant while in the CV lab. Participants spent an average of 69.5 min (SD = 34.8) in the CV lab; this time did not differ among groups [Kruskal–Wallis ANOVA \( \chi^2(df = 2) = 2.3, p = .31 \)].

Randomization to compression method resulted in 99 participants assigned to Femostop®, 108 to C-clamp, and 99 to manual (additional information provided by the author expanding this article can be found on the editor’s Web site at http://www.nursing-research-editor.com). There were no differences by compression method for body surface area or baseline diastolic blood pressure (see editor’s Web site for additional information). There was a difference in baseline systolic blood pressure (see editor’s Web site for additional information), which was determined not to be clinically meaningful to the primary study aims and was not considered as a covariate. Mean time to hemostasis was 33.4 min (SD = 15.6) with a range of 15–175 min, which was found to be non-normal. Mean time to hemostasis by device was as follows: Femostop® 40.2 min (SD = 23.2), C-clamp 32.6 min (SD = 9.8), and manual 27.5 min (SD = 6.3). Time to hemostasis was found to be significantly different among the compression methods [Kruskal–Wallis ANOVA \( \chi^2(df = 2) = 53.2, p < .0001 \)] and was used as a covariate in subsequent analyses.

Discomfort
Nurses administer medications to patients to alleviate discomfort associated with prolonged bed rest and femoral artery compression. Prior to sheath removal, participants received oral acetaminophen with codeine (Tyleol #3; 33%), intravenous morphine sulfate (9%), oral propoxyphene napsylate with acetaminophen (Darvocet; 7%), subcutaneous (SQ) lidocaine (79%), and oral hydromorphone (Dilaudid; 51%), which were controlled for in subsequent analyses.

Subjects reported a range of discomfort of 0–10 over the study period for each compression method (additional information provided by the author expanding this article can be found on the editor’s Web site at http://www.nursing-research-editor.com). Overall, subjects reported mild discomfort associated with sheath removal and compression of the femoral vasculature at each assessment point, with a mean from 0.5 to 2.0 (see editor’s Web site for additional information). Discomfort increased slightly from baseline 1 min after compression was applied for all subjects, with those randomized to C-clamp reporting slightly greater discomfort. Discomfort ratings decreased to baseline levels 10 min after compression was applied with further decrease to minimal levels 10 min after compression was released.

Controlling for medications and time to hemostasis, there were no significant differences in discomfort ratings by compression method (Table 1). The only significant
Effects of Three Groin Compression Methods

**TABLE 1. Generalized Estimating Equation (GEE) Models by Outcomes and Compression Method**

<table>
<thead>
<tr>
<th>Outcome and Variables in the Model</th>
<th>Z statistic</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discomfort (over 4 time points)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SQ lidocaine</td>
<td>1.05</td>
<td>.29</td>
</tr>
<tr>
<td>Hydromorphone (Dilaudid)</td>
<td>-0.75</td>
<td>.45</td>
</tr>
<tr>
<td>Morphine sulfate</td>
<td>-0.64</td>
<td>.52</td>
</tr>
<tr>
<td>Acetaminophen/codeine (Tylenol #3)</td>
<td>-4.4</td>
<td>.0001</td>
</tr>
<tr>
<td>Acetaminophen/propoxyphene naplate (Darvocet)</td>
<td>-0.44</td>
<td>.66</td>
</tr>
<tr>
<td>Time to hemostasis</td>
<td>-1.0</td>
<td>.32</td>
</tr>
<tr>
<td>Groin compression method*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Femostop®</td>
<td>0.21</td>
<td>.83</td>
</tr>
<tr>
<td>C-clamp</td>
<td>0.15</td>
<td>.88</td>
</tr>
<tr>
<td>Distress (over 4 time points)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midazolam (Versed)</td>
<td>0.17</td>
<td>.87</td>
</tr>
<tr>
<td>Time to hemostasis</td>
<td>0.32</td>
<td>.75</td>
</tr>
<tr>
<td>Groin compression method*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Femostop®</td>
<td>-0.74</td>
<td>.46</td>
</tr>
<tr>
<td>C-clamp</td>
<td>-1.2</td>
<td>.23</td>
</tr>
<tr>
<td>Oozing (over three time points)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline oozing</td>
<td>-3.1</td>
<td>.0019</td>
</tr>
<tr>
<td>Eptifibatide (Integrilin)</td>
<td>-0.27</td>
<td>.79</td>
</tr>
<tr>
<td>Heparin</td>
<td>1.6</td>
<td>.11</td>
</tr>
<tr>
<td>Clopidogrel (Plavix)</td>
<td>0.18</td>
<td>.85</td>
</tr>
<tr>
<td>Femostop®, ≤30 min to hemostasis*</td>
<td>-1.7</td>
<td>.08</td>
</tr>
<tr>
<td>Femostop®, &gt;30 min to hemostasis*</td>
<td>-2.3</td>
<td>.02</td>
</tr>
<tr>
<td>C-Clamp, ≤30 min to hemostasis</td>
<td>-2.2</td>
<td>.03</td>
</tr>
<tr>
<td>C-Clamp, &gt;30 min to hemostasis</td>
<td>-1.6</td>
<td>.12</td>
</tr>
<tr>
<td>Manual, ≤30 min to hemostasis</td>
<td>-2.3</td>
<td>.02</td>
</tr>
<tr>
<td>Hematoma (over three time points)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline hematoma</td>
<td>-9.4</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Eptifibatide (Integrilin)</td>
<td>-1.3</td>
<td>.18</td>
</tr>
<tr>
<td>Heparin</td>
<td>-1.3</td>
<td>.18</td>
</tr>
<tr>
<td>Clopidogrel (Plavix)</td>
<td>-0.56</td>
<td>.57</td>
</tr>
<tr>
<td>Time to hemostasis</td>
<td>0.30</td>
<td>.77</td>
</tr>
<tr>
<td>Groin compression method*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Femostop®</td>
<td>-0.34</td>
<td>.74</td>
</tr>
<tr>
<td>C-clamp</td>
<td>-0.47</td>
<td>.64</td>
</tr>
<tr>
<td>Ecchymosis (over three time points)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline ecchymosis</td>
<td>-10.1</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Eptifibatide (Integrilin)</td>
<td>-2.9</td>
<td>.004</td>
</tr>
<tr>
<td>Heparin</td>
<td>1.4</td>
<td>.15</td>
</tr>
<tr>
<td>Clopidogrel (Plavix)</td>
<td>-1.2</td>
<td>.22</td>
</tr>
<tr>
<td>Time to hemostasis</td>
<td>1.8</td>
<td>.07</td>
</tr>
</tbody>
</table>

**TABLE 1. (continued).**

<table>
<thead>
<tr>
<th>Outcome and Variables in the Model</th>
<th>Z statistic</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groin compression method*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Femostop®</td>
<td>-0.87</td>
<td>.38</td>
</tr>
<tr>
<td>C-clamp</td>
<td>1.1</td>
<td>.29</td>
</tr>
</tbody>
</table>

Note. Coding: All outcome measures are dichotomized as present/absent at each time point that include baseline oozing, ecchymosis, and hematoma; all medications dichotomized as yes/no.

Time to hemostasis is interval, except where noted.

*Manual, hemostasis >30 min is reference method.

Contributor to discomfort ratings was for acetaminophen with codeine (Tylenol #3). Participants not receiving acetaminophen with codeine (Tylenol #3) were less likely to report any discomfort.

**Distress**

Per protocol for the unit, nurses usually administer IV midazolam (Versed) to patients to allay distress before sheath removal; 94% of participants received this medication. As with discomfort ratings, subjects reported a range of distress of 0–10 over the study period (see editor’s Web site for additional information). Overall, subjects report minimal distress at each of the assessment points, with a mean of 0.3–1.1 (see editor’s Web site for additional information). Distress decreased slightly from baseline 1 min after sheath removal, except for manual compression which resulted in very slight increased distress. Distress ratings continued to trend downward at the remaining two assessment points.

There were no significant differences in distress ratings by compression method (Table 1). Even though or because a majority of participants (94%) received midazolam (Versed), this medication did not have an influence on distress ratings; effect could not be measured because almost every patient received the medication.

**Vascular Complications**

Incidence of VCs prior to femoral sheath removal (Table 2) included 58 (19%) participants with oozing, 57 (18%) with ecchymosis, and 59 (19.3%) with hematoma, all of which were controlled for in subsequent analyses. A number of participants received antplatelet medications while in the CV lab (heparin; 92%) or prior to sheath removal [eptifibatide (Integrilin, 38%); clopidogrel (Plavix, 3%);] all were used as covariates in subsequent analyses.

Incidence of pulsatile mass after sheath removal was limited to one subject and was not considered in further analysis. Overall, presence of VCs in participants ranged from 6 (2%) to 112 (37%; Table 2). Incidence of oozing trended downward after sheath removal across all groin compression methods. Ecchymosis increased at each assessment point after sheath removal for all compression methods and tended to be more frequent in those participants randomized to C-clamp. Hematoma increased slightly for the Femostop® and C-clamp groups between...
immediate femoral sheath removal and the 12-hr assessment period. Presence of hematoma in the manual group tended to decrease over the assessment period. Incidence of oozing by compression method was not significant. The baseline presence of oozing prior to sheath removal and time to hemostasis significantly influenced the incidence of oozing (Tables 1 and 2). Compared to the manual compression method with >30 min to hemostasis, Femostop® was found to contribute to oozing significantly less when participants had a greater median time to hemostasis. C-clamp was found to contribute to oozing significantly less when time to hemostasis was less than the median of 30 min.

Incidence of ecchymosis by compression method was not significant (Tables 1 and 2). Baseline presence of ecchymosis and receiving eptifibatide (Integrilin) were significant contributors to ecchymosis. Participants without baseline ecchymosis who did not receive eptifibatide (Integrilin) were significantly less likely to have ecchymosis after femoral sheath removal.

Incidence of hematoma by compression method was not significant. Baseline presence of hematoma prior to sheath removal was the only significant contributor to hematoma (Tables 1 and 2). Participants without baseline hematoma were significantly less likely to have a hematoma after femoral sheath removal.

Discussion

Assessment and management of VCs and symptoms are important nursing care responsibilities for PCIP patients. The aims of this study were to determine if one of three compression methods (Femostop®, C-clamp, or manual) after femoral sheath removal following PCIP causes less patient discomfort and distress and causes fewer VCs. Findings from this randomized trial did not result in one method of compression to cause less discomfort or distress. Participants in this study reported mean discomfort of 2.0 overall, whereas others have found sheath removal to be slightly more uncomfortable (2.65; Puntillo et al., 2001); manual compression was the least painful (1.9) and Femostop® was the most painful (3.1; Lehmann et al., 1999). Mean distress ratings from participants in this study are comparable to other investigators (Puntillo et al., 2001).

While this study is the only known investigation to compare all three methods of compression simultaneously and include participants who had received or were receiving anticoagulant or antiplatelet medications, there was no difference in VCs by compression method. Patients who did not have a VC prior to femoral sheath removal were less likely to have a VC after the procedure. This might be due to the specific use of Femostop® to reduce oozing when time to hemostasis was longer or shorter than the median time of 30 min. Using Femostop® contributed to less oozing when time to hemostasis was prolonged and using a C-clamp contributed to less oozing when time to hemostasis was shorter.

The most frequently occurring VC was ecchymosis (37%) followed by hematoma (20%) and oozing (19%; Table 2). These findings are similar to those reported by others for ecchymosis 29–61% (Lehmann et al., 1999; Pracyk et al., 1998) and hematoma 10–36% (Lehmann et al., 1999; Pracyk et al., 1998). Incidence of oozing was slightly higher in this study (19%) than reported by others (3–16%; Lehmann et al., 1999; Pracyk et al., 1998).

<table>
<thead>
<tr>
<th>VCs by compression method</th>
<th>Number present at baseline (%)</th>
<th>Number present immediately after sheath removal (%)</th>
<th>Number present 12 hr after sheath removal (%)</th>
<th>Number present 24 hr after sheath removal (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oozing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Femostop®, n = 99</td>
<td>20 (7%)</td>
<td>9 (3%)</td>
<td>6 (2%)</td>
<td>2 (0.6%)</td>
</tr>
<tr>
<td>C-Clamp, n = 108</td>
<td>25 (8%)</td>
<td>11 (3.5%)</td>
<td>8 (2.4%)</td>
<td>2 (0.6%)</td>
</tr>
<tr>
<td>Manual, n = 99</td>
<td>13 (4%)</td>
<td>12 (4%)</td>
<td>5 (1.6%)</td>
<td>2 (0.6%)</td>
</tr>
<tr>
<td>Total, N = 306</td>
<td>58 (19%)</td>
<td>32 (10.5%)</td>
<td>19 (6%)</td>
<td>6 (2%)</td>
</tr>
<tr>
<td>Ecchymosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Femostop®, n = 99</td>
<td>18 (6%)</td>
<td>19 (6%)</td>
<td>32 (11%)</td>
<td>32 (11%)</td>
</tr>
<tr>
<td>C-Clamp, n = 108</td>
<td>17 (5%)</td>
<td>21 (7%)</td>
<td>41 (13.3%)</td>
<td>42 (14%)</td>
</tr>
<tr>
<td>Manual, n = 99</td>
<td>22 (7%)</td>
<td>24 (8%)</td>
<td>39 (12.7%)</td>
<td>33 (11%)</td>
</tr>
<tr>
<td>Total, N = 306</td>
<td>57 (18%)</td>
<td>64 (21%)</td>
<td>112 (37%)</td>
<td>107 (36%)</td>
</tr>
<tr>
<td>Hematoma</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Femostop®, n = 99</td>
<td>21 (7%)</td>
<td>19 (6%)</td>
<td>20 (7%)</td>
<td>13 (4%)</td>
</tr>
<tr>
<td>C-Clamp, n = 108</td>
<td>18 (5.5%)</td>
<td>15 (5%)</td>
<td>22 (7%)</td>
<td>16 (5%)</td>
</tr>
<tr>
<td>Manual, n = 99</td>
<td>20 (6.5%)</td>
<td>20 (6%)</td>
<td>18 (6%)</td>
<td>16 (5%)</td>
</tr>
<tr>
<td>Total, N = 306</td>
<td>59 (19%)</td>
<td>54 (17%)</td>
<td>60 (20%)</td>
<td>45 (14%)</td>
</tr>
</tbody>
</table>

*TABLE 2. Presence of Vascular Complications (VCs) by Groin Compression Method (N = 306)*
perhaps due to the inclusion of patients receiving anti-coagulant and antiplatelet medications.

The lack of significant findings may be attributed to several factors. First, power analysis calculations were based on a sample size of 306 subjects needed to detect a difference of 60% in discomfort ratings by compression method, and a 20–40% difference among compression methods for VCs. Though the target sample size was recruited, discomfort and distress ratings and VCs did not vary greatly by compression methods. Furthermore, study RNs were all experienced in the care of patients post-PCIP, which may have contributed to no difference in VCs or symptoms by compression method.

**Limitations**

A convenience sample of predominantly older White men was recruited from one site in the urban Midwest. This limits the generalizability of findings to patients who are younger, ethnically diverse, and of the female gender.

Costs associated with compression method and nursing time to achieve hemostasis were not assessed in this study. Further work is needed to detail costs associated with this common nursing procedure, particularly focusing on the cost of nursing time and the development of repetitive use syndrome in nurses’ extremities with manual compression of the femoral artery.

Another limitation concerns the primary data analysis. Due to the data being non-normal, nonparametric analyses were required that did not accommodate specific discomfort and distress ratings nor specific doses of medications, only if a participant had the symptom present/absent or received a medication. Symptom ratings and dosages of medications may be significant contributors and requires consideration in future studies.

**Directions for Future Research and Nursing Practice**

Participants in this study who received pain medications for symptoms not associated with actual femoral artery compression continued to report discomfort, indicating current medications are not effective. Research is needed to test other pain medication regimens, including timing of medication administration, which has been found to be a major issue in pain relief for chest tube removal (Punttillo & Ley, 2004).

Receipt of midazolam (Versed) prior to sheath removal was not found to influence distress ratings statistically. Further study is necessary specifically to test the effectiveness of midazolam on distress ratings and whether non-pharmacologic interventions, such as imagery or music, could be useful adjuncts to allay distress in those patients who require intervention.

While none of the three compression methods were found to be superior in reducing VCs, other factors may need to be considered. Demographic and clinical characteristics that might be predictive of complications are older age, female gender, unstable angina, and hypertension (Waksman et al., 1995) and warrant further examination.

Findings from this study did not indicate that instillation of SQ lidocaine prior to sheath removal was effective in alleviating groin area discomfort. Two previous investigations also concluded SQ lidocaine infiltration is not effective in reducing groin discomfort (Bowden & Worrey, 1995; Wadas & Hill, 1998). Nursing care should focus on administering the appropriate pain medication for the patient’s pain quality.

The presence of VCs prior to sheath removal was found to significantly influence the presence of complications after hemostasis was achieved. Nurses should be aware of those factors that may increase the intensity of patient care by planning interventions accordingly, particularly when a VC is present before sheath removal and antiplatelet medications have been administered.

Results from this study do not indicate one method of compression to be superior for minimizing VCs or causing less discomfort and distress. Significant contributors include the presence of VCs prior to sheath removal and receipt of certain antiplatelet medications. Further study is needed to discern demographic and patient characteristics that may be significant contributors to VCs. Nurses should pay extra attention when providing care to patients who have a pre-existing VC and have received antiplatelet medications.

**References**


Application of the CuSum Technique to Evaluate Changes in Recruitment Strategies

Patrick McNees • Karen Hassey Dow • Victoria Wochna Loerzel

Background: While lagging subject enrollment in longitudinal clinical trials is a complex problem, the best recruitment strategy has not been established. Cumulative summation (CuSum) is a statistical process control procedure often applied in quality improvement efforts to detect trend shifts in highly variable serial data.

Objectives: To describe changes in efforts to increase referrals and enrollment in a longitudinal quality-of-life breast cancer study, determine effects of changes in referral strategies on enrollment using a novel application of CuSum, and discuss implications of CuSum as a tool for prospectively managing the subject recruitment process.

Method: Ten referrals and eight enrollments per month for a total of 31 months were estimated to meet study subject accrual requirements in the clinical trial. The estimates were used as standards in performing CuSum calculations. CuSum was applied to monthly referral and enrollment data and trend graphs were generated. Alterations in recruitment tactics and strategies were evaluated as to whether changes in trend occasioned such alterations. Unplanned changes in trend were noted.

Results: While monthly data were highly variable, an average of 8.42 referrals and 5.92 enrollments were realized during Months 1–12. Based on these figures, projected accrual for 31 months would have enrolled only 184 subjects, 66 subjects short of target. CuSum illustrated this trend. Subsequent shifts in enrollment trends were shown with improvements in referral.

Discussion: Indications for use of CuSum include (a) earlier detection of enrollment trend shifts, and (b) earlier discrimination between effective and ineffective recruitment. Thus, CuSum has implications for both evaluating the effects of planned and unplanned process changes and for managing the recruitment process.

Key Words: accrual • CuSum • recruitment

Lagging subject enrollment in longitudinal clinical trials can be a taxing problem for investigators. Recruitment issues are complicated by controllable and uncontrollable factors (Butterfield, Yates, Rogers, & Healow; 2003; Cooley et al. 2003; Cox & McGarry, 2003; Grap & Munro, 2003; Neumark, Stommel, Given, & Given, 2001). Over a decade ago, Ashery and McAuliffe (1992) reviewed nine clinical trials and found that subject recruitment was the most common implementation problem. However, there continues to be little empirical evidence differentiating the utility of various recruitment strategies or establishing the effects of changes in recruitment efforts over time. Additionally, there is a paucity of both analytic and management tools to help researchers detect and address lagging enrollment earlier rather than later in the research process. These problems pose threats to the research enterprise by disrupting the data collection plan, extending the overall recruitment and enrollment timeline, hampering statistical analysis, and increasing research-related costs. Ultimately, the validity of individual studies can be compromised due to an inadequate sample size, thus reducing the impact on both research and practice (Rushforth, 2005).

There are several barriers to recruitment and enrollment in clinical trials. These barriers include clinicians’ attitudes (Cox & McGarry, 2003), inability to give consent in the emergency and critical care settings (Grap & Munro, 2003), multiple clinical sites (Cooley et al., 2003), minority and ethnic differences (Brown, Fouad, Basen-Engquist, & Tortolero-Luna, 2000), and advancing age (Neumark et al., 2001).

Attempts by Fairhurst and Dowrick (1996) to enroll subjects in a randomized trial to evaluate the effectiveness of counseling in the management of minor psychiatric morbidity were confounded by the attitudes and behavior of general practitioners. They found that practitioners lacked motivation for study involvement, had ethical doubts about randomization, and perceived a lack of a viable noncounseling intervention.

Patrick McNees, PhD, is President and Chief Scientist, Applied Health Science, Inc., Orlando, Florida; Karen Hassey Dow, PhD, RN, FAAN, is Professor and Kabli Endowed Chair in Oncology Nursing, School of Nursing; and Victoria Wochna Loerzel, RN, MSN, OCN, is Project Director, Breast Cancer Education Intervention and Instructor, University of Central Florida, Orlando, Florida.
In formulating a research plan or developing a successful grant proposal, investigators are well trained and experienced in identifying an adequate sample size through a power analysis. Investigators often describe the potential clinical sites available for subject recruitment, discuss the target number for enrollment, and can produce a well-detailed targeted or planned enrollment table including the gender, ethnic, and racial categories of the planned sample size. While there are a multitude of controllable and uncontrollable issues that affect enrollment, potential problems of lagging recruitment are sometimes overlooked until it is clear that an adequate sample may be jeopardized. Frequently, even less attention is paid to the proportion of referrals that are projected to enroll. The number of investigators who actively monitor referral patterns over time is unclear. One common strategy to overcome the results of such limitations is to increase the number of clinical sites available for recruitment. Yet, additional efforts to recruit these sites, navigate through additional institutional review board (IRB) procedures, and cultivate the requisite research collaboration can be extraordinarily time consuming. Too often there still remains an inadequate sample size.

The purposes of this study are as follows: (a) to describe changes in recruitment, referral, and enrollment strategies in a longitudinal quality-of-life clinical trial in breast cancer; (b) to determine effects of changes in referral strategies on subject enrollment in a longitudinal clinical trial using a novel application of cumulative summation (CuSum); and (c) to discuss implications of CuSum as a tool for aiding in the management of the recruitment, referral, and enrollment process. The use of CuSum in nursing research and an empirical evaluation of applying CuSum to the subject recruitment, referral, and enrollment process are described.

**Early Referral and Enrollment Patterns in the Breast Cancer Education Intervention Study**

The Breast Cancer Education Intervention (BCEI) is a randomized longitudinal trial testing the effects of targeted psychoeducational interventions in 250 recently diagnosed early stage breast cancer survivors. Subject accrual estimates for the 4-year study were based on assumptions of an average of 10 referrals and eight enrollments per month over a 31-month enrollment period.

Similar to other longitudinal clinical trials, and reflected in Figure 1, the research team experienced a wide variation in monthly subject accrual. The investigators monitored monthly referral and enrollment numbers. During Months 1–6, there were a total of 54 referrals and 35 enrollments, 13 short of target enrollment. In response to these lower-than-expected numbers, the research team employed additional recruitment methods often cited as alternate means of increasing enrollment. One effort was to expand recruitment sites to additional community-based cancer centers. Other efforts included attending and providing information at community health fairs. These efforts were time intensive and drained financial resources and were not particularly effective. By Month 12 of data collection, 101 subjects were referred and 71 subjects were enrolled, still 25 short of target enrollment. Thus, increasing recruitment sites resulted in insufficient enrollment to satisfy study requirements. The research team redirected and refocused efforts to increasing and strengthening referrals from the original recruitment site. The general strategy was to do whatever was necessary to remove barriers to referral and sustain high visibility within the referring institution. A more specific delineation of these efforts follows.

Toward the end of the first year of subject enrollment (Months 10–12), the research staff reorganized efforts to concentrate on maximizing its current collaborative research relationship with the original referring cancer center. The collaborative research team identified several fixable barriers. Perhaps the greatest barrier to enrollment involved the specific time at which potential subjects were informed of their eligibility. Referring staff assumed that potential subjects should be informed of the study at the time they became eligible, after conclusion of treatment. While such practice is common and appropriate in some clinical trials, informing potential subjects of their eligibility after conclusion of treatment proved problematic and resulted in lower than expected enrollment. There are a couple of possible reasons for the effect. First, eligible subjects were approached at a time when they no longer wanted to be reminded that they had cancer and wanted to “get on with their lives” (Dow, 1994; Dow, Ferrell, Leigh, Ly, & Gulasekaram, 1996; Dow, Ferrell, Haberman, & Eaton, 1999; Ganz et al., 2004). Second, eligible subjects were identified and approached by a research nurse with whom they were unfamiliar, thus making them less likely to participate. To address these issues, the clinical staff at the cancer center agreed to modify the recruitment approach. They moved away from considering the BCEI project a traditional clinical trial and attempted to identify and approach potential subjects at diagnosis or while still on treatment. Approaching them before treatment ended improved the prospects for entry into the study.

A second barrier concerned the referring research nurses at the cancer center. Documenting subject eligibility was a time-consuming process. The process entailed verifying diagnosis and pathology reports and photocopying the physicians’ progress reports and sending them to the research office. The research team streamlined this process by developing a quick eligibility checklist that could be faxed to the research office, thus decreasing almost an hour of paperwork to a few minutes.

Finally, lack of regular communication and contact between the teams of researchers at the university and clinical staff at the cancer center resulted in a disconnection between the two staffs. One result was that the cancer center clinical staff simply forgot to refer patients for the study. The group used several strategies to manage this problem. First, the research team contacted the referring physicians to get renewed commitment. Physicians and nurses were reminded of the eligibility criteria and the process of referral and were encouraged to discuss the study with their patients. Second, the research team devised and implemented a feedback loop. Specifically, the team generated monthly referral and enrollment reports along with physician rankings (i.e., how each physician compared to
other physicians in terms of number of referrals) and distributed them to the referring physicians. Third, research team members began weekly visits to the cancer center in an attempt to increase visibility, provide an additional stimulus for referrals, and improve communication. During these weekly visits, research team members also made rounds to identify eligible patients and to update referral estimates. These efforts resulted in a dramatic increase in enrollment during Months 13–19 (108 referrals and 75 subjects enrolled). While efforts were successful in achieving more than adequate referral and enrollment during this 7-month time period, cumulative study referrals and enrollments continued to lag behind projected total numbers. At Month 25, the investigators discussed whether they could have determined the downward trends in referral and enrollment sooner and conducted a retrospective CuSum analysis to evaluate whether the changes in recruitment influenced referral and enrollment trends.

**Literature Review**

**Research as a Process**

The research process has fallen largely outside the realm of systematic evaluation. Discussions and propositions for improving the process are set forth only rarely. Bowman, Wyman, and Peters (2002) suggest that the formulation and use of an operations manual is one way to improve the research process. Yet researchers often view recruitment activities, referrals, enrollment, and attrition as a set of discrete but related events and outcomes. However, if the events are viewed as representing a continuous process that optimally should produce stable patterns and predictable outcomes, then the paradigm resembles one that has been recognized by manufacturers for decades. The paradigm is one of quality control and quality improvement.

**Statistical Process Control**

In 1924, while at Western Electric Company, Walter Shewhart (1931) first applied statistical techniques to manufacturing in attempts to identify and influence assignable sources of variability while not responding to chance sources of variability. These visual charts with statistical control limits were referred to as “control charts.” In the 8 decades since then, it is logical to assume that other tools might have been developed and applied successfully that could be used to improve research processes.

During the 1930s, Shewhart’s (1931) statistical process control charts began to be adopted as standards for industrial applications and were advocated for use in World War II (WWII) to improve production. Shewhart’s work began to draw the attention of several individuals who would later revolutionize manufacturing processes and quality control efforts. One such individual was W. Edwards Deming (1943, 1982). During WWII, another individual consolidated a body of work started in the 1930s. Abraham Wald (1947) provided the mathematical and statistical bases for a family of techniques based on sequential analysis.

One of these techniques, CuSum, is part of this family of sequential analysis techniques used regularly in statistical process control for quality improvement (Page, 1954). While Shewhart-type control charts are used to detect large magnitude changes (SD > 0.5), CuSum has been effectively to detect small trend shifts in highly variable serial data. CuSum plots can be applied to serially collected data. However, unlike more typical considerations of sample distributions or linear rate analysis, rather than simple deviation from the mean, CuSum is used to show the trend and directionality of linear data. CuSum charts plot the cumulative sum of the deviations between each data point and a reference value (Bishop & Nix, 1993). In brief, CuSum shows changes in the trend rather than focusing on the distance between plotted points and the centerline.

**CuSum Applications in Healthcare**

CuSum was described fully by Page in 1954 and by 1960 was being reported in the medical literature (Armitage, 1960). CuSum has been used since in a variety of clinical
healthcare quality control situations including assessment of physician performance (Lim, Soraya, Ding, & Morad, 2002), surgical trainee competence in aptitude and psychomotor skills (Van Rij et al., 1995), surgical competence with parotid surgery (Sharp, Cozens, & Robinson, 2003), trainee competence in anesthesia procedures (Bolsin & Colson, 2000), trends in use of laparoscopic cholecystectomy (Bartlett & Parry, 2001), learning curves involved with endovascular abdominal aortic aneurysm repair (Forbes, DeRose, Kribs, & Harris, 2004), and quality control analysis of single subject data (Shehab & Schlegel, 2000).

CuSum has been applied also in a wide variety of clinical research problems, including plotting temperature to evaluate the effectiveness of antimicrobial treatment of neutropenic patients (Kinsey, Giles, & Holton, 1989), analyzing circadian blood pressure profiles in hypertensive patients (Stanton, Cox, Atkins, O’Malley, & O’Brien, 1992), evaluating the effectiveness of a wound protocol in reducing postoperative skin breakdown after joint arthroplasty (Chang & McLean, 2002), evaluating whether computer-assisted surgery can improve technical outcomes in total knee replacement surgery (Nizard et al., 2004), detecting random shifts and trends in tumor markers after curative resection among patients with breast cancer (Schlain, Lavin, & Hayden, 1993), and evaluating the sensitivity of serial serum squamous cell carcinoma antigen for monitoring and detecting recurrence in patients with head and neck cancer (D’Amico, Snyderman, Wagner, & Nerella, 1995).

CuSum has been demonstrated recently to be functional in early and rapid detection of infectious diseases (O’Brien & Christie, 1997) and salmonella outbreaks (Hutwagner, Maloney, Bean, Slutsker, & Martin, 1997). The potential role of CuSum application and utilization in disease surveillance has been discussed (Yang et al., 1997).

Methods

A retrospective evaluation was performed of the recruitment, referral, and enrollment processes using CuSum technique to help detect shift in trends in the data. Monthly referrals and enrollments were entered into a Microsoft Excel spreadsheet. Then the data were checked independently for accuracy. The following CuSum formula was used:

\[ C_s = \sum_{i=1}^{n} (e_i - T_0) \]

where \( e \) = each event or datum for the slope (i.e., number referrals or enrollments each month) and \( T \) equals the target or standard (i.e., approximately 10 referrals and 8 enrollments).

In setting up the formulae for Excel calculations, the following pseudocode was used for referrals. For the first cell, the CuSum point = [(actual referrals for the Month 1) – (250/0.8/31)]. For the second cell, the CuSum point = CuSum point 1 + [(actual referrals for Month 2) – (250/0.8/31)], and so on. Enrollment CuSum points were calculated in a similar fashion except (250/31) was used for the monthly target enrollment. For example, assume referrals were in Column A and that the first data point was in cell A1. The formula for the first CuSum point in B1 would be as follows: +A1 – (250/0.8/31), the CuSum point in B2 would be +B1 + (A2 – (250/0.8/31), and so on. In these examples, 250 reflects the target accruals. Because it was anticipated that for every 10 referrals there would be 8 enrollees, the accruals are adjusted by dividing by 0.8, thus allowing for a 20% refusal rate. Finally, the calculation includes a divisor of 31 to adjust the data to a per month basis.

The analysis will be described in four phases. Phase I consists of the numerous attempts to increase the number of referral sites and ways to directly recruit subjects from the community. Phase II was focused on the strategic and tactical changes made with the original cancer center to assure that necessary systems were in place to improve referral. In Phase III, there was unplanned reallocation of resources and personnel changes at the cancer center. During Phase IV, a new research nurse was assigned by the cancer center to oversee recruitment.

![FIGURE 2. CuSum plot of enrollment for Months 1 through 12.](image-url)
Results

To illustrate the functionality of CuSum in evaluating the effects of recruitment efforts on subject enrollment and accrual, consider the enrollment data for Months 1–12 (Phase I) illustrated in Figure 1. Despite attempts to expand and diversify recruitment efforts, the data reflect slower accrual than had been anticipated. While data were highly variable from month to month, the overall rates were problematic with an average of 8.42 referrals and 5.92 enrollments per month in Months 1–12. Based on these data, only 184 subjects would have been enrolled by Month 31, 66 subjects short of target. The CuSum plot (Figure 2) clearly illustrated not only the fact that the accrual rate did not meet projection, but also that efforts to expand and diversify recruitment did not result in a sustained change in trend and therefore did not have the desired overall effect.

During Months 1–10, the research team began to monitor referral patterns and enrollment. As with enrollments, monthly referral patterns were highly variable with the exception that overall rates were too low and the pattern was difficult to interpret. When Months 1–12 referrals were plotted using CuSum, a much clearer pattern emerged (Figure 3). Special attention is drawn to the change in the referral trend during Month 12. While this change could represent a change in trend, similar changes occurred during Months 6 and 8. More data would be needed to confirm this change.

Figure 4 reflects the monthly referrals and enrollment for the 31 months of subject enrollment. Again, the apparent highly variable data make interpretation difficult. When the same data are plotted in CuSum, several interesting points emerge (Figure 5). On Month 12 there was a change in the referral trend; however, a 1-month change had been noted previously. Additionally, at least two data points are required to confirm a trend shift. Thus, more data were required to confirm an actual change. Referring to referral data for Month 13, the trend is disrupted still from the first 11 months, confirming the shift in trend.

When CuSum plots for referrals and enrollment are presented together, trend changes in referrals appear to be reliable predictors of subsequent trend changes in enrollment.

Discussion

CuSum is used effectively to show the troublesome Phase I trend in referrals and enrollment. CuSum also appears to be a tool that can be used to detect trend changes in variable referral and enrollment patterns. The ability to detect such changes is important in evaluating the effectiveness of recruitment efforts as well as suggesting earlier when recruitment tactics or strategies should be altered. Consider the first 14 months of the study, that is, Phase I and the first few months in Phase II. During the first 11 months, there is a negative trend largely unaffected by efforts to improve enrollment. However, between Months 11 and 12, there was a change in trend for referrals that was confirmed by subsequent data and followed by a trend shift for enrollments. Changes in recruitment efforts directed toward direct education and information (e.g., health fairs) and efforts to broaden the referral base to additional referral sites had little effect. However, the effects of focusing on the physician–nurse referral processes, providing routine feedback, and maintaining a persistent and functional presence resulted in about 8 months of uninterrupted improvement in referrals and enrollment.

It is important to note that the CuSum did not result in improvements. The CuSum is a tool that assists in detecting shifts in trends. The technique is particularly useful for time series data that are highly variable. As such, CuSum may be used to determine more quickly the effects of practices, unintended alterations in practice, and deliberate practice modifications.

Also CuSum can be used to detect unplanned changes in referral and enrollment patterns. The consequences of such an unplanned change can be noted in the Phase III trend shift in referrals beginning in Month 20 and the subsequent shift in enrollments beginning in Month 21. This change in trend
occurred when the cancer center research staff responsibilities were reorganized and the assigned research nurse at the cancer center decreased involvement in the BCEI study and eventually left the cancer center. When a new highly supportive research nurse was assigned to the BCEI study in Phase IV, the changes in trend were dramatically apparent. It appears that the processes and tactics employed at the referral site were necessary but insufficient to maintain referral and enrollment targets without a committed individual or champion from the cancer center.

The researchers described an application of CuSum in nursing research. While CuSum appears to have evaluation and management utility for the research process, it is only one tool. Other tools are more appropriate in certain circumstances. For example, a traditional Shewhart-type control chart is an even simpler aid for detection large changes (changes between 0.5 and 2.5 SD) to signal when a process is out of control. Additionally, the recruitment process is only one of several processes related to the conduct of longitudinal trials that has received little attention. Methods for managing data that result in better security, higher reliability, and increased integrity have been discussed but have fallen beyond the scope of systematic investigation and evaluation.

Future research should be directed toward the identification of other tools that systematically support and improve the research process, and methodologies that allow investigators to approach improving research processes with the same rigor they try to bring to core study designs for longitudinal trials. It is through such work that research processes can be moved in a more scientifically credible fashion and improved incrementally.

**Figure 4.** Monthly referral and enrollment for Months 1 through 31. Solid line reflects enrolled and broken line reflects referral. --- Referral; --- Enrolled.

**Figure 5.** CuSum plot of referral and enrollment by study phase for Months 1 through 31. Solid line reflects enrolled and broken line reflects referral. --- Referral; --- Enrolled.
References


Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
Methods

Multilevel Modeling of a Clustered Continuous Outcome

Nurses’ Work Hours and Burnout

Sunhee Park ▪ Eileen T. Lake

Background: Multilevel models were designed to analyze data generated from a nested structure (e.g., nurses within hospitals) because conventional linear regression models underestimate standard errors and, in turn, overestimate test statistics.

Objectives: To introduce 2 types of multilevel models, the random intercept model and the random coefficient model, to describe the correlation among observations within a cluster, and to demonstrate how to identify the superior model.

Method: The conceptual and mathematical bases for the 2 multilevel model types are presented. Intraclass correlation is defined and assessment of model fit is detailed. An empirical example is presented in which average work hours per week and burnout are analyzed using data from 4,320 staff nurses clustered in 19 hospitals.

Results: Average work hours were positively associated with nurse burnout. The multilevel models corrected the problem of underestimated standard errors in conventional linear regression models. Graphs displaying the hospital-level differences illustrated the 2 multilevel model types. Although the multilevel models corrected the underestimation of standard errors, the results did not differ substantively for the conventional or the 2 multilevel models. The intraclass correlation coefficient was .044, indicating that the extent of shared variance among nurses in a hospital was low. The random intercept model fit the data better than did the random coefficient model.

Conclusions: Multilevel models provide a more accurate and comprehensive description of relationships in clustered data than do conventional models, by correcting underestimated standard errors, by estimating components of variance at several levels, and by estimating cluster-specific intercepts and slopes.

Key Words: clustered data · hierarchical structure · multilevel models

Observations of data collected from a hierarchical structure can be correlated with one another. Examples of hierarchical data are (a) nurses nested within hospitals and (b) repeated measurements of the same person. In these instances, nurses and repeated measurements are at Level 1 and hospitals and persons are at Level 2 (Kreft & De Leeuw, 1998). With clustered data, estimated standard errors in a conventional regression are smaller than actual standard errors due to failure to account for the correlated structure of observations. This underestimation of standard errors causes an important problem, the increase of a Type I error, which implies a higher possibility of concluding that obtained results are significant, even though they may not be so (Reise & Duan, 1999).

Hence, the incorporation of correlated responses is crucial to obtain accurate results in the analysis. Multilevel models were developed to correct the dependency of observations within a cluster. Multilevel models have been popular among researchers in social and behavioral sciences because most data are generated from hierarchical structures (Kreft & De Leeuw, 1998; Raudenbush & Bryk, 2002).

Compared to the above disciplines, few studies in nursing have attempted to fix problems originating from correlated observations. The nursing researchers who have used multilevel models have focused on the effects of nurse staffing and nursing unit characteristics on patients’...
and nurses’ outcomes (Cho, Ketefian, Barkauskas, & Smith, 2003; Mark, Harless, McCue, & Xu, 2004; Mark, Salver, & Wan, 2000, 2003; McGillis Hall et al., 2003). Although Wu (1995) introduced multilevel models and Cho (2003) addressed the importance of multilevel modeling in nursing outcomes research, they did not detail various types of multilevel models nor did their examples contrast conventional regression results with the results of multilevel models.

The purpose of this article is to explore in depth the concept and use of multilevel models. First, multilevel models are introduced and two types are explained: the random intercept model and the random coefficient model. Second, the intraclass correlation (i.e., the amount of dependency among observations) and the method of choosing the better model in the comparison of two models are described. These concepts and methods are illustrated in a multilevel analysis of nurse survey data with a clustered continuous outcome. Nurse burnout is regressed on nurses’ average work hours per week. The results of the two multilevel models are compared with those of a conventional linear regression model. Finally, graphs are provided that illustrate the sophisticated statistical treatment of clustered data that multilevel models offer.

Types of Multilevel Models

The conventional linear regression in Equation 1 presumes all observations are independent.

\[ y_i = \alpha + \beta x_i + e_i \quad (1) \]

where \( \alpha \) is the intercept, \( \beta \) is the effect of the covariate \( x_i \), and \( e_i \) is the error term. The independence of observations signifies that the error terms for observations are unrelated to one another. The error terms, \( e_i \), are also assumed to be normally distributed with the mean of 0 and the variance of \( \sigma^2 \) (Moore & McCabe, 2003).

If the values of \( y_i \) are correlated as a consequence of the clustering of observations at a second level, the error terms among observations become correlated because the error term represents omitted factors in a model or measurement error in a dependent variable. The correlation of the error terms in clustered data violates the assumption in a conventional linear regression (Allison, 1999a). When this assumption is violated, the conventional regression does not perform well in estimating standard errors of coefficients. The correlated responses yield smaller standard errors of coefficients. Consequently, there is a higher possibility of obtaining falsely significant results (Allison, 1999b). Thus, the application of a multilevel model is appropriate due to the fact that multilevel models take into consideration the correlated structure of observations in the estimation of standard errors.

Multilevel models are divided into the random intercept model and the random coefficient model, based on how to treat an intercept and covariates. The random intercept model only considers the intercept as a random component, whereas the random coefficient model treats covariates as well as the intercept as random variables. It is also possible to treat only covariates as random variables. By treating the intercept and covariates as random components, the variation and effects of clusters can be examined (Kreft & De Leeuw, 1998). Contrary to the conventional multilevel models, the conventional linear regression model does not consider the plausibility of similarity among observations in a cluster, and simply postulates that all the error terms are not related to one another. Here, the random intercept model and the random coefficient model are presented with an example of one covariate in a two-level structure.

Two-Level Random Intercept Model

The random intercept model (Equation 2) is the simplest type of multilevel model. Only the intercept is allowed to differ across clusters.

\[ y_{ij} = (\alpha + \beta x_{ij}) + (u_j + e_{ij}) \quad (2) \]

where \( y_{ij} \) is the dependent variable for observation \( i \) in cluster \( j \), \( \alpha \) is the intercept, \( \beta \) is the effect of the covariate \( x_{ij} \), \( u_j \) is the Level-2 random effect, and \( e_{ij} \) is the Level-1 random effect.

Equation 2 can be expressed in two equations based on a hierarchical structure. When equations 3 and 4 are combined, Equation 2 is obtained:

Level 1: \( y_{ij} = \beta_{xij} + e_{ij} \)

Level 2: \( \beta_{xij} = \alpha + u_j \) \quad (4)

Unlike Equation 1, the random intercept model has two components of the intercept in Equation 4. \( \alpha \) is the average intercept across all observations and all clusters, and \( u_j \) is the amount by which the intercept of cluster \( j \) deviates from the average \( \alpha \). In Equation 2, \( e_{ij} \) is the error term associated with observation \( i \) in cluster \( j \) and \( u_j \) is the error term of all the observations in cluster \( j \). Therefore, among all observations in the same cluster, the error term \( u_j \) is the same, but the error term \( e_{ij} \) could be different. In the random intercept model, the intercept \( \beta_{xij} \) can vary across clusters, but the coefficient \( \beta \) of the covariate \( x_{ij} \) does not differ across clusters. Consequently, the intercept across clusters could be different but the slope across clusters would be the same (Healy, 2001). The random intercept model can also be considered to have two parts, fixed effects and random effects. In Equation 2, \( \alpha + \beta x_{ij} \) does not change, regardless of observations and clusters, but \( u_j + e_{ij} \) could change according to clusters and observations. Because of these characteristics, the former and the latter can be called fixed effects and random effects, respectively (Kreft & De Leeuw, 1998).

The random effects at the cluster level, \( u_j \), are assumed to have the mean of 0 and the variance of \( \sigma^2 \). Also, the random effect in cluster \( j \) should not be correlated with either the random effects at the observation level, \( e_{ij} \), or the random effect of other clusters. The assumptions related to \( e_{ij} \) are the same as those in the conventional linear regression model (Hox, 2002).

Two-Level Random Coefficient Model

In the random coefficient model (Equation 5), the slope and the intercept are allowed to vary across clusters.

\[ y_{ij} = (\alpha + bx_{ij}) + (x_{ij}u_j + e_{ij}) \quad (5) \]

where \( y_{ij} \) is the dependent variable for observation \( i \) in cluster \( j \), \( \alpha \) is the intercept, \( b \) is the effect of the covariate
The random coefficient model in Equation 5 differs from the random intercept model (Equation 2) because this model has two components of the effect of the covariate in Equation 8. \( b \) can be interpreted as the average of the covariate effects across clusters, and \( v_i \) is the amount of deviation from the average covariate effect, \( b \), in cluster \( j \). Hence, in the random coefficient model, each cluster can have a different intercept and slope. In Equation 5, all clusters have the same values for the part \( \alpha + bx \), but each cluster could have different values for the part \( v_i x + u_i + e_{ij} \). Thus, like the random intercept model, the random coefficient model is separated into fixed and random elements (Kreft & De Leeuw, 1998).

The random effect of a covariate at the cluster level, \( v_p \), is assumed to have a mean of 0 and variance \( \sigma^2_v \). Also, covariance between the random effects \( u_i \) and \( v_i \) is possible. Except for the assumption regarding the term \( v_i \), and the allowance of covariance between \( u_i \) and \( v_i \), all other assumptions in the random coefficient model are same as those in the random intercept model (Guo & Zhao, 2000).

Just as the conventional linear regression model sometimes includes multiple covariates, the random coefficient model with multiple covariates is also feasible. In addition, it is possible to treat several covariates as random variables at the same time. This makes it plausible to investigate the effects of several covariates at Level 2. However, there is a trade-off in using additional covariates. With additional covariates, more assumptions must be fulfilled and more parameters should be estimated (Hox, 2002). The assumptions and parameter estimation require more complex computations and therefore it will take longer to obtain results.

### Characteristics of Multilevel Models

Multilevel models have three features that differ from conventional linear regression models. First, as noted earlier, the use of conventional regression for clustered data results in the underestimation of standard errors, because this model does not consider the similarity of responses among observations within the same cluster. Smaller standard errors affect test statistics, which are more likely to be statistically significant. Multilevel models resolve this problem by including random components of cluster effects in the statistical model. The consideration of random effects at the cluster level in the multilevel model makes it possible to estimate correct standard errors. By dividing the total variance in the dependent variable into between-cluster and within-cluster parts, the variability of random effects across clusters and the importance of clusters can also be evaluated (Teachman & Crowder, 2002; Wu, 1995).

Second, both observation-level and cluster-level covariates can be included in multilevel models. Owing to this feature, cross-level interactions (i.e., the relationship between observation-level and cluster-level covariates) can be examined. A cross-level interaction is useful for research questions about whether cluster-level characteristics moderate individual-level relationships. For example, country-level characteristics such as gross national product might moderate the individual-level relationship between women's educational attainment and fertility (Guo & Zhao, 2000; Hox & Kreft, 1994; Teachman & Crowder, 2002).

Third, aggregation bias might occur in the conventional linear regression model, but can be eliminated in multilevel models. Aggregation bias implies that the results from aggregated variables analyzed at the cluster level may be different from those at the original observation level (Kreft & De Leeuw, 1998). Wu (1995) indicates that multilevel models separate the estimated effects in the covariates into different levels, which can be interpreted as individual-level effects (i.e., within a cluster) and cluster-level effects (i.e., across clusters), respectively.

### Intraclass Correlation

Intraclass correlation refers to correlation among observations within a cluster. The intraclass correlation coefficient (ICC) measures this degree of correlation. An ICC can be determined from an intercept-only model (i.e., a multilevel model with no covariates):

\[
\text{Level 1: } y_{ij} = \beta_0 + \epsilon_{ij} \\
\text{Level 2: } \beta_0 = \alpha + u_i \\
\beta_j = b + v_j
\]

By combining equations 9 and 10, Equation 11, the intercept-only model, is produced. The intercept-only model does not account for any variance in the dependent variable. It only separates the variances of the dependent variable into two parts; that is, the variance of clusters, \( \sigma^2_u \), and the variance of observations at Level 1, \( \sigma^2_e \). The ICC, written as the symbol \( \rho \), can be computed on the basis of these two variance components (Equation 12). \( \rho \) ranges from 0 to 1.

\[
\rho = \frac{\sigma^2_u}{\sigma^2_u + \sigma^2_e}
\]

If all the observations are independent of one another, the ICC equals 0. At the other extreme, if all the responses from observations in all clusters are exactly the same, the ICC equals 1. A nonzero ICC implies that the observations are not independent. If observations are highly correlated, the variance of observations at Level 1, \( \sigma^2_e \), becomes smaller. In turn, the denominator in Equation 12 becomes smaller, which means that ICC becomes larger (Hox, 2002).

The ICC is interpreted in four ways. First, the ICC represents the degree of common environments that observations share. The ICC would increase if observations in the same cluster were under more similar environments and, as a result, if the responses of...
observations became more alike. The second interpretation of the ICC is the proportion of total variance (i.e., cluster plus individual variance) that is attributed to the cluster level, which is self-explanatory from Equation 12. Therefore, as the relative variance of the clusters increases, the less likely you are to assume that the groups are similar. The third interpretation of the ICC is the degree of homogeneity of Level 1 within Level 2 (i.e., the quantity of similarity among observations at the cluster level). If observations were not correlated, they would not affect one another nor would they be similar at all (i.e., no homogeneity). The last interpretation of the ICC is the anticipated correlation between two observations that are randomly chosen from the same cluster (e.g., correlation of two nurses within the same hospital) (Hox, 2002; Kreft & De Leeuw, 1998).

Choosing the Best Model

Several multilevel models can be tested (e.g., a two-level random intercept model and a two-level random coefficient model) depending on the researcher's purposes. The estimation of more than a two-level model is also possible, and there is no restriction regarding the number of random covariates (Hox, 2002; Kreft & De Leeuw, 1998). The choice of a relevant model is an important step, and it should be based on the necessity of parsimony in a model. Parsimony means that models should be as simple as possible (Hox, 2002).

The loglikelihoods of the two models are compared to select the model that best fits the data, if one model is a special case of the other model (Hox, 2002; Kreft & De Leeuw, 1998). The choice of a relevant model is an important step, and it should be based on the necessity of parsimony in a model. Parsimony means that models should be as simple as possible (Hox, 2002).

To compare the fit of two models, calculate a \( \chi^2 \) statistic as the positive difference of 2\( \times \) their respective loglikelihoods. The positive difference of 2\( \times \)loglikelihood has a \( \chi^2 \) distribution with degrees of freedom obtained from the difference of the number of parameters to be estimated in the two models. If the statistic is not statistically significant, this suggests that a more complicated model may not be necessary (Hox, 2002). For instance, if a random intercept model and a random coefficient model are compared and if the \( \chi^2 \) statistic is not significant, this means that a random intercept model, which contains fewer parameters, is acceptable. The comparison of loglikelihoods can also be used to investigate the significance of a series of random coefficient variables in selecting a better model between a model that contains these random variables and a model that does not.

Empirical Example

Sample

The data were obtained from New Zealand hospital staff nurses in an extension to an international study that investigated the effects of both nurse staffing and nursing practice environments on hospital patient outcomes in five countries (Aiken, Clarke, & Sloane, 2002). The sample was 4,320 staff nurses in 19 acute general public hospitals in New Zealand. The number of nurses in each hospital ranged from 28 to 547. Nurses answered questions about their current job, burnout, and job satisfaction.

Model for Testing

A continuous dependent variable from a two-level structure was used to explore the random intercept model and the random coefficient model. The model for the analysis was:

Random Intercept Model:

\[
Burnout_{ij} = \beta_0 + \beta \times \text{Average work hours}_{ij} + e_{ij}
\]  

(13)

Random Coefficient Model:

\[
Burnout_{ij} = \beta_0 + \beta_j \times \text{Average work hours}_{ij} + e_{ij}
\]  

(14)

Equation 13 is the random intercept model and Equation 14 is the random coefficient model. In Equation 13, only the intercept \( \beta_0 \) could be different across clusters and the slope \( \beta \) was considered fixed, whereas in Equation 14, both the intercept \( \beta_0 \) and slope \( \beta \) could be different at the second level. nurse burnout was used as the dependent variable and measured using nine items of “emotional exhaustion," a subscale of the Maslach Burnout Inventory. Each item score ranged from 0 to 6. The nurses’ burnout score was calculated as the sum of the nine item scores, and could vary from 0 to 54. A higher score of emotional exhaustion indicated that the amount of burnout was more intense (Maslach, Jackson, & Leiter, 1996). The average work hours per week in the past year was used as the covariate.

Results

For all respondents, the mean burnout score was 21.9 (SD = 11.2), which corresponds to an average level of burnout on the basis of published norms for a sample of nurses and physicians (Maslach et al., 1996). Across hospitals, the mean burnout score ranged from 16.5 (SD = 12.2) to 25.9 (SD = 12.0). The mean for work hours per week was 34.0, with an SD of 10.2. Across hospitals, the mean work hours per week varied from 30.2 (SD = 9.0) to 35.7 (SD = 11.2).

Underestimation of Standard Errors in the Conventional Linear Regression Model

As stated earlier, standard errors from clustered data may be underestimated in a conventional linear regression model (Hox, 2002). As expected, the standard errors in the conventional linear regression model were smaller than or equal to those in the random intercept model and the random coefficient model, although the difference...
of standard errors between these models was small (Table 1).

**Random Intercept Model**

The random intercept model treated only the intercept as a random component and estimated the variance of the intercept at the hospital level. The coefficients of the intercept (i.e., the degree of burnout when work hours are 0) and the covariate (i.e., the average effect of work hours on the degree of burnout) at the nurse level were also assessed. The intercept was 16.4, which stood for the average burnout score for all nurses in all hospitals, when work hours were 0. The coefficient estimate of work hours was 0.15, which implied that with a 1-hour increase of work per week, the average nurse’s burnout score increased by 0.15. The intercept and the slope of work hours were statistically significant (p < .0001).

Unlike the conventional linear regression model that had the variance at the nurse level only, the random intercept model had two variances of the intercept: variance at the nurse level and variance at the hospital level. The nurse-level and hospital-level variances were separately estimated as 118.78 and 4.02, respectively. Even though the variance of the intercept at the hospital level was fairly small compared with the variance at the nurse level, the hospital-level variance of the intercept was statistically significant (p < .001). This result signified that the amount of burnout was significantly different across hospitals when the value of work hours was 0. The intercepts of the conventional linear regression model and the random intercept model were interpreted differently. The intercept of 16.4 as a fixed variable in the conventional regression implied that all nurses in all hospitals had the same burnout score when work hours were 0, and this score is considered unchanged. However, the intercept of 16.4 as a random variable in the random intercept model meant that although the mean intercept in all nurses across all hospitals was 16.4, the intercept could differ in particular hospitals. Based on the value of the intercept and the variance at the hospital level, the 95% predicted interval that contained the intercepts in the hospitals was calculated from the formula

\[ a \pm 2 \times \sqrt{\sigma_u^2} \]  

(Hox, 2002). The 95% predicted interval of the intercept (i.e., burnout score when work hours per week is 0) across hospitals was from 12.4 to 20.4 in this example.

The existence of variance at the hospital level implied that the intercept could differ across hospitals. At the same time, the random intercept model did not allow the slope of work hours to vary across hospitals. Figure 1 illustrates the results of the random intercept model: the same slope of work hours, but different intercepts across hospitals. Each line represents the relationship between the average work hours per week and nurses’ burnout in one hospital. All lines show that the number of work hours was positively associated with the amount of nurses’ burnout, but the levels of burnout were not exactly the same when work hours were 0 (i.e., the intercept). The lines are parallel, which means that the slope, that is, the effect of the work hours on burnout, is the same across hospitals.

**Random Coefficient Model**

The parameter estimates of the intercept and the slope of work hours in the random coefficient model were almost the same as those in the random intercept model (Table 1). The difference between these models was that the random coefficient model allowed variability of the coefficient of the covariate at the hospital level (i.e., the effect of work hours on the degree of burnout at Level 2). The variability of work hours at the hospital level meant that the slope was no longer the same across hospitals. The variance of the slope of work hours at the hospital level was .002, and the covariance of the intercept and the slope was −0.071.

### TABLE 1. Parameter Estimates and Standard Errors in a Conventional Linear Regression Model and Multilevel Models

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Conventional Linear Regression Model</th>
<th>Intercept-only Model</th>
<th>Random Intercept Model</th>
<th>Random Coefficient Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>SE</td>
<td>Coefficient</td>
<td>SE</td>
</tr>
<tr>
<td>Fixed part</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>16.40**</td>
<td>0.602</td>
<td>21.39**</td>
<td>0.250</td>
</tr>
<tr>
<td>Work hours</td>
<td>0.17**</td>
<td>0.017</td>
<td>0.15**</td>
<td>0.017</td>
</tr>
<tr>
<td>Random part</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma_b^2 )</td>
<td>121.46**</td>
<td>2.655</td>
<td>118.78**</td>
<td>2.636</td>
</tr>
<tr>
<td>( \sigma_u^2 )</td>
<td>5.64**</td>
<td>1.315</td>
<td>4.02*</td>
<td>1.157</td>
</tr>
<tr>
<td>( \sigma_v^2 )</td>
<td></td>
<td></td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>( \sigma_w^2 )</td>
<td></td>
<td></td>
<td>−0.071</td>
<td>0.087</td>
</tr>
<tr>
<td>Loglikelihood</td>
<td>−15514.78</td>
<td></td>
<td>−15513.64</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** \( \sigma_b^2 \): Variance at the nurse level; \( \sigma_u^2 \): Variance of the intercept at the hospital level; \( \sigma_v^2 \): Variance of the coefficient of work hours (i.e., the slope of work hours) at the hospital level; \( \sigma_w^2 \): Covariance between the intercept and the coefficient at the hospital level; S.E.: Standard Error.

\*p < .001. **p < .0001.
This negative correlation would indicate that at the hospital level, higher values of baseline burnout (at theoretical zero for work hours) would be associated with a relatively smaller effect of work hours on burnout. However, the hospital-level random effects (i.e., the variability of the intercept and the slope of work hours and the covariance between them at the hospital level) were not statistically significant. This finding implied that the slope of work hours across hospitals did not differ notably. Figure 2 illustrates the results of the random coefficient model. Here, the slopes of the lines, representing the effect of work hours on burnout in each hospital, are not parallel and the intercepts differ across hospitals.

**Intraclass Correlation and Likelihood Ratio Test**

The intraclass correlation was calculated from the results of the intercept-only model in Table 1. The intercept-only model had no covariates. Hence, this model divided the variance of burnout into two components: nurse level and hospital level (Hox, 2002). The intraclass correlation based on these two variance components (see Equation 12) was $5.64/(121.46 + 5.64) = .044$. Thus, the variance at the hospital level accounted for 4.4% of the total variance of burnout.

The likelihood ratio test was conducted to see whether the random intercept model or the random coefficient model better fit the data. The positive difference of $2 \times \loglikelihood$, 2.28, in two models followed a $\chi^2$ distribution with one degree of freedom. The $p$ value of the test statistic was .13, which demonstrated that the simpler model, the random intercept model, fit the data better. This result also indicated that the different slopes of work hours across hospitals could be eliminated from the model.

**Discussion**

In this presentation, a model with one covariate was introduced to illustrate and to compare the results of the random intercept model versus the random coefficient model. The multilevel models demonstrated the correction of the underestimated standard errors in the conventional linear regression model. Standard errors in the conventional regression were not substantially underestimated, however, which happens when observations within clusters are not highly correlated. The small ICC of .044 in the example demonstrates the low degree of correlation among observations within a cluster. The comparison of the loglikelihood between the two models showed that the simpler model, the random intercept model, was more appropriate.

Although the ICC was quite low in the empirical example, it was consistent with existing literature. In a review of organizational research literature, James (1982) reported a median ICC of .12 for measures of organizational climate. A lower ICC for nurse burnout as compared with that for organizational climate would seem reasonable. Greater perceptual agreement would be expected within organizations about an organizational phenomenon such as climate than about an individual outcome such as burnout. Indeed, in a study of 632 nurses on 54 nursing units in four hospitals, Forbes and Taunton (1994) reported a substantially lower ICC of .003 at the hospital level and .04 at the nursing unit level for a different job outcome, job enjoyment. That study demonstrated hospital-level ICCs of .03 to .07 and nursing unit ICCs of .03 to .23 for three organizational phenomena: nurse manager structuring expectations, nurse manager consideration, and control over practice. These hospital versus nursing unit differences in the ICC reflect greater perceptual agreement at the workgroup level than at the institutional level. Data analyzed by Hughes and Anderson (1994) from 223 registered nurses, licensed vocational
nurses, and nurses aides on a skilled nursing unit in each of 14 nursing homes yielded ICCs of .03 and .02, respectively, for two organizational phenomena: decentralization and participation in decision making. Thus, ICCs for both organizational and individual phenomena appear from the literature to be low in general.

When observations are more highly correlated, the difference of standard errors between the conventional regression model and the multilevel models becomes greater. In the empirical example, there was a minimal difference of standard errors between conventional and multilevel models, consistent with the low ICC. Moreover, the substantive findings (i.e., the effect of work hours on burnout) did not differ between models. In instances where the ICC is low, the use of methods that seek to address the underestimated error term may not be necessary.

The application of multilevel models for clustered data has attractive features: (a) the correction of underestimation of standard errors, (b) the examination of the cross-level interaction, (c) the elimination of concerns about aggregation bias, and (d) the estimation of the variability of coefficients at the cluster level. Moreover, it is feasible to choose the better model through the likelihood ratio test and to estimate the amount of intraclass correlation (Kreft & De Leeuw, 1998; Teachman & Crowder, 2002; Wu, 1995).

Even if multilevel models have the above strengths, they should be used only after prudent consideration of the following factors. First, the process of maximum likelihood, which determines parameter estimates in multilevel models, is computationally intense because iterations continue until the best parameters are determined (Hox, 2002). As the number of parameters to be estimated increases, it takes longer to obtain those estimates. This means that the random coefficient model is more computationally intense than the random intercept model (Wong & Mason, 1985). Second, when more parameters are estimated using the available data, precision might be lost because to estimate additional parameters, new assumptions regarding those parameters are needed (Raudenbush & Sampson, 1999). These two factors suggest how important it is to choose a parsimonious model for testing to yield efficiency and precision.

It is essential to take into account the contexts in which observations are collected because of the possibility of correlated responses among observations. Multilevel models can yield reliable results by reflecting a hierarchical structure, thereby correcting the underestimation of standard errors. These models can also unveil the importance of the hierarchical structure in the relationships between factors of interest. Therefore, the application of multilevel models should be considered in the analysis of clustered data.

Accepted for publication March 1, 2005.
This research was supported by a grant from the National Institute of Nursing Research (NR004513). The principal investigator is Linda H. Aiken, PhD, RN, FAAN, FRCN, Director of the Center for Health Outcomes and Policy Research at the University of Pennsylvania.

The authors thank Dr Linda H. Aiken for permission to use her data. Ms Park also thanks Dr Paul D. Allison, Professor, Department of Sociology, University of Pennsylvania, for his sincere encouragement.

Corresponding author: Sunhee Park, MPH, RN, School of Nursing, University of Pennsylvania, 420 Guardian Dr, Philadelphia, PA 19104 (e-mail: sunheep3@nursing.upenn.edu).

References


Methods

Evaluating Content Validity for Children’s Self-Report Instruments Using Children as Content Experts

Janet L. Stewart • Mary R. Lynn • Merle H. Mishel

Background: The development and evaluation of instruments to index cognitive and emotional processes from the perspectives of children is a priority for pediatric nursing research.

Objective: To describe the procedures used in employing children as content validity experts in the development of a self-report instrument.

Methods: Following published recommendations for moving from qualitative research to quantitative measurement with adults and for maximizing content validity in self-report instruments, six children aged 8–16 years undergoing treatment for cancer constituted the panel of content experts for review of a measure of children’s illness-related uncertainty derived from qualitative interviews. Children were provided with an explanation of the project, an explanation of their role as experts, and explicit instructions on how to evaluate the representativeness of individual items and the total scale.

Results: Generally, the children performed the review tasks effectively, although two children (ages 8 and 16 years) had initial difficulty in going beyond their own experience when considering the relevance of individual items. Twenty items were deemed acceptable by at least five out of the six children and two additional items were revised based on their input.

Discussion: Employing children as content validity experts adds a critical dimension to establishing psychometrically sound measures for studying the processes affecting the health of children and families.

Key Words: child • content expert • content validity • measurement

Children’s under-representation in self-report instruments severely hampers understanding of many important processes affecting the health of children and families (Deatrick, Faux, & Moore, 1993; Johnson, 1991; Young, Yoshida, Williams, Bombardier, & Wright, 1995). Proxy reports of children’s emotional and social behaviors, particularly by parents and other adults, are frequently used to index feeling states such as depression and anxiety, although these reports have been shown to correlate poorly with children’s self-reports (Varni, Katz, Colegrove, & Dolgin, 1995). Therefore, the development and evaluation of instruments to index children’s cognitive and emotional processes is a priority for research into their psychological adjustment to illness (Coster & Haley, 1992; deLeeuw, Borgers, & Smits, 2004; Johnson, 1991).

One promising method for developing valid children’s self-report measures is using qualitative data generated by children to define and index the content domain of interest. Several nurse researchers have detailed the procedures used to move from qualitative research to quantitative measurement with adults and in doing so have established legitimacy and precedence for the practice (Fleury, 1993; Imle & Atwood, 1988; Sofaer, 2002; Tilden, Nelson, & May, 1990). One of the recommended steps is employing individuals who served as informants in the original qualitative study to evaluate the content validity of the resulting instrument (Fleury, 1993; Tilden et al., 1990). However, applying these procedures when developing measures to be used with school-aged children and adolescents requires consideration of the cognitive...
capacity of children to engage in the complex task of evaluating whether an instrument is comprehensive and cohesive.

The purpose of this presentation is to describe the steps taken to adapt the content validity review process to engage children aged 8–16 years as experts. The self-report instrument under review was developed from interviews with children undergoing treatment for cancer (Stewart, 2003) and designed to measure uncertainty in children and adolescents.

Established procedures for maximizing content validity for self-report instruments include domain identification, item generation, and judgment and quantification (DeVellis, 2003; Grant & Davis, 1997; Lynn, 1986). The principles of qualitative data collection and analysis are consistent with the first two steps; that is, assuring that the concept is well understood (domain identification) and carefully explicating (item generation). The procedures for using qualitative data from child informants to generate items for a quantitative instrument are no different than those used with adult informants. The coding, categorization, and synthesis that serve as analytic tools for deriving qualitative meaning provide the dimensional structure for an instrument, and individual informants’ data provide the basis of the instrument items representing each dimension (Fleury, 1993). As is true with all instrument development from qualitative data, staying true to children’s qualitative data in item development ensures that the instrument items will reflect the language and experiences of the children targeted for eventual employment of the instrument.

The final step in establishing content validity, judgment and quantification, has traditionally been implemented by inviting review by a panel of experts who judge each item and the total scale for its representation of the conceptual domain (Grant & Davis, 1997; Lynn, 1986). When an instrument is derived from qualitative data, using subjects from the qualitative study to serve as content experts has been proposed as a means of assuring that the instrument reflects the meaning of the phenomenon being measured (Fleury, 1993; Imle & Artwood, 1988; Tilden et al., 1990). No published reports were found in which qualitative informants who were children were asked to serve as content experts. However, several authors reported using focus groups made up of children to evaluate instruments undergoing development or revision (Christie, French, Sowden, & West, 1993; de Leeuw et al., 2004; Hockenberry-Eaton et al., 1998). In these reports, children were described as being particularly helpful in providing feedback about what is missing from a scale, especially if there are perspectives or contexts not addressed in the existing items. They also gave very valuable feedback about how well they understand items and instructions.

Regardless of the age of those who serve as content validity experts, a critical step in ensuring a participant’s ability to fulfill the expert role is the provision of clear expectations for the task. Grant and Davis (1997) and Lynn (1986) recommend that reviewers be provided with the conceptual definition of the construct being measured, given explicit instructions for how to rate items and the total scale, and encouraged to suggest revisions and additions based on experience. The desirable attributes of lay experts are that they understand the conceptual definitions and criteria for evaluating the instrument and its items, and that they are able to see beyond their own experiences to the collective representation of the concept. However, even adult lay experts may have trouble understanding the terminology and abstractions typically used by professionals, or seeing beyond their own experiences to evaluate the collective experience of others (Fleury, 1993; Sandelowski, 1993). Therefore, extra care must be taken in providing child experts with developmentally appropriate materials that will enable them to contribute meaningfully to content validity review.

**Approach**

The primary goal of adapting the content validity review process was to accommodate the developing comprehension and language abilities of children in order to assure that they understood not only the instrument items but also what was being asked of them as experts. Children were oriented to the role of the content validity expert, making each step of the review process explicit and providing them with an evaluative framework that used familiar language in rating the validity of the items and total scale.

As required by the institutional review board, children asked to serve as experts read and signed an assent form that explained the project and their role in it. Language in the assent form was assessed to be at a third grade reading level and the instructions were reviewed orally with each child expert to uncover any questions or misunderstandings. Also provided was a written explanation of the task to child experts in the form of a cover letter that detailed the elements of the process, as recommended by Grant and Davis (1997) and Lynn (1986). Excerpts from the cover letter illustrating the explanations provided can be seen in Table 1.

Two challenges were encountered in preparing children to serve as content experts: explaining the nature of survey research to younger children and helping children understand the role of being a critical reviewer. The cover letter included an explanation of survey research and the reason for developing the instrument (Table 1). In the written instructions included in the cover letter and in the pretask verbal discussion, children were encouraged to view the expert role as consulting rather than merely responding to items from their own perspective.

The child experts were then provided with a scoring sheet with boxes to mark their judgment for each item. To monitor children’s thoughts and reactions, the investigator engaged in the task with each child expert: reading the items, recording their responses and their comments, and providing clarification as necessary. If a child responded that an item was “okay,” the investigator asked him or her to think about what might make it a better item. If a child responded that an item was “bad,” the investigator asked for an explanation and whether the child thought the item could be made better. Each child was also asked if there were any additional
TABLE 1. Excerpts From the Cover Letter Explaining Content Validity Review for Children

<table>
<thead>
<tr>
<th>Content Validity</th>
<th>Excerpts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose of the project</td>
<td>&quot;I want to ask lots of kids with cancer about uncertainty. The best way to ask a lot of kids the same questions is to use a SURVEY. A survey is a written form with statements, called ITEMS. There are boxes after each item for kids to mark their answers. A survey is not a test, so there are not any right or wrong answers. I'm using the things kids told me about having cancer to make a survey about uncertainty.&quot;</td>
</tr>
<tr>
<td>Conceptual definition</td>
<td>&quot;Uncertainty means the kinds of things kids can't always be sure about. For example, they may not be sure where their cancer came from. Or, they may not be sure how bad the side effects from treatment might be.&quot;</td>
</tr>
<tr>
<td>The expert role</td>
<td>&quot;I am asking you to help me with my survey because you are an EXPERT! I've come up with items for my survey. Now I'd like your EXPERT OPINION about whether you think each item should be in my survey for kids. I want my survey for kids to be the best it can be, and that's why I'm asking the experts!&quot;</td>
</tr>
<tr>
<td>Instructions for item review</td>
<td>&quot;I'd like you to think back about the kinds of things you've felt and thought since getting cancer. Also, based on what you know about having cancer, I'd like you to think about what a kid your age might think or feel about getting cancer. Then, tell me if you think each item should be in the survey.&quot;</td>
</tr>
<tr>
<td>Scoring system</td>
<td>&quot;A GOOD ITEM is one that describes something a kid with cancer might be unsure or uncertain about. A good item also uses words that are easy to understand. An OKAY ITEM is one that describes something a kid with cancer might be uncertain about, but isn't easy to understand. A BAD ITEM is one that doesn't have anything to do with kids and uncertainty. A bad item could also be one that doesn't make any sense.&quot;</td>
</tr>
</tbody>
</table>

Sample

Six children undergoing treatment for cancer agreed to serve as content validity experts; one 16-year-old boy declined to participate, citing time demands. Three of the six children were adolescents aged 14–16 years who had participated in the earlier qualitative study. Three younger children (ages 8–11 years) who had not participated in the earlier study were recruited in order to broaden the age range of the panel to parallel that of the target population for the instrument. The first author met with each child individually, one child in her home and the remaining children in the outpatient setting.

Results

The scale consisted of 22 items derived from discrete data bits from the qualitative study of children's uncertainty in the context of cancer treatment (Stewart, 2003). Using the Simple Measure of Gobbledygook formula (Lynn, 1989), the reading level of the 22 items was determined to be third grade, which was deemed the desirable minimum reading level because the target population for the instrument is children aged 8–16 years. As it generally is recommended that instrument reading level be kept below sixth grade even for adults (Davis, Crouch, Wills, Miller, & Abdehou, 1990), the third-grade level was considered appropriate for an instrument that spanned late childhood and early adolescence. Children were asked to adopt a broad perspective in judging each item and not to limit their judgment of the items to their own experiences. However, two children (one 8-year-old and one 16-year-old) consistently rated items as “bad” if the item was inconsistent with their experience; that is, if they did not feel that way or had not experienced a particular feeling or thought. Both children were allowed to complete the survey from their own perspective and then they were asked to complete it a second time, answering for an imaginary child with cancer or one they knew who was not exactly like them. With this second review, these two children were able to broaden their perspective and consider more items as relevant to the content domain while identifying items they thought were confusing or irrelevant.

Following Lynn’s (1986) recommendation that at least five out of six experts should agree on an item’s relevance to include it in the scale, items were retained if five of the children agreed that an item was “good” or “okay.” For 20 of the original 22 items, at least five of the six child experts agreed that the item was “good” or “okay.” Children’s comments were used to revise several items to improve clarity, including the two items that were rated as “bad” by more than one child expert. For example, the item stem “[I am unsure] if things will go the same each time I get treatment” was revised to “[I am unsure] if there will be any surprises when I get a treatment.” None of the children identified additional content areas or items that should be added to the instrument.

Because no published support was found for employing children as content experts, a more conventional review panel of clinical experts was convened also. Four nurse clinicians with expertise in childhood cancer were asked to review the 22 items that had been revised after the child experts’ input. The nurses were asked to rate each item on a 4-point Likert scale for relevance to the content domain, to suggest revisions that would improve the relevance, and to identify content not indexed by scale items. According to Lynn’s (1986) criteria for content validity, all four clinical experts were required to rate each item as 4 (relevant as presented) or 3 (relevant with minor revisions) to
include that item in the scale. These criteria were met for each of the 22 items and all items were retained for pilot testing. None of the clinical experts suggested additional content.

**Discussion**

Experts from the fields of both qualitative inquiry and instrument development have noted the value of grounding measurement in the expertise of knowledgeable informants—those who have experienced the phenomenon of interest (Mishel, 1997; Morse, 1991). Qualitative research has been recognized not only as offering a strong foundation for quantitative measurement, but also as a particularly effective approach to representing the unique perspective of children (Broome, Hellier, Wilson, Dale, & Glanville, 1988; Docherty & Sandelowski, 1999; Siegal, 1997). With careful consideration of the maturational limitations on conceptual and linguistic complexity, children can provide abundant information and rich detail about their daily lives as well as exceptional experiences, such as illness. When children are used as content experts, the researcher must make sure they understand the expert role, the concept being explored, and the interpersonal context of participating in a research study (Deatrick et al., 1993; Faux, Walsh, & Deatrick, 1988; Garbarino & Stott, 1992).

In maximizing the potential for using qualitative research to generate instruments that are grounded in the experience of children and thereby retain conceptual validity, it is critical to address children’s capabilities and limitations at each step of the research process. There have been few guidelines for adapting established psychometric procedures (such as content validity review) to the special case of child measures. Therefore, the following were used here: the procedures outlined by Grant and Davis (1997) and Lynn (1986) for preparing content experts; the importance of making the evaluation process clear to lay participants as suggested by Fleury (1993); and the first two authors’ experiences as pediatric nurses to develop instructions and procedures that children could use to provide critical input about the instrument under development.

The primary challenge encountered was assisting children to go beyond their individual experiences to consider the relevance of individual items to the content domain. This did not appear to be simply a function of age, as one of the two participants most challenged by the task was 16 years old, and others have noted similar challenges with adult lay reviewers (e.g., Fleury, 1993; Sandelowski, 1993). Therefore, it may be desirable when using children as content experts to consider a larger pool of potential experts than the minimum needed and to replace experts who are unable to access the broader sphere of content necessary to complete item and instrument assessment.

As experts, the children were asked to use their experiences, observations, and insights to consider what the collective population of children with cancer might think and feel. The children worked diligently and thoughtfully in considering the relevance of individual items and provided very helpful suggestions for revising items to improve relevance and clarity. No major revisions were made based on the children’s review (no items were deleted or added), which supports that deriving quantitative data is effective in establishing a critical content domain and generating items that effectively and comprehensively index it. Although a conventional panel of clinical experts was employed in addition to the child informants given the novelty of this approach, it is recommended that developmentally sensitive, systematic review by child experts is sufficient in determining content validity for a child-focused, self-report instrument.

Conducting research with children provides real challenges and also unique opportunities to represent their lives and experiences from their own perspectives. Employing children as content validity experts is consistent with both the qualitative principle of member checking (inviting qualitative informants to verify a researcher’s interpretations to enhance the credibility of the results) and the crucial step of establishing content validity in instrument development. It thereby adds a critical dimension to establishing psychometrically sound measures for studying the processes affecting the health of children and families.

**References**


Background: For the majority of grandparents, rarely are their perceptions of family assessed, acknowledged, or viewed in the context of potential effect on individual and family health.

Objective: To develop and test an instrument to measure grandparent perceptions of family.

Method: An inductively derived semantic differential instrument comprising 112 items, 8 stimuli, and 27 adjective pairs within 2 domains and framed by the Becoming a Grandparent theoretical model was content validated by an expert panel, revised, and evaluated on a convenience sample of 306 community-dwelling grandparents, aged 41–92 years (M = 64, SD = 9.8), 72% female, 93% U.S. born, 62% college educated, 55% economically comfortable, and 71% living with a spouse or partner. Assessment procedures included internal consistency reliability, item and scale correlation coefficients, and principal components factor analysis with varimax rotation to evaluate construct validity, structure, and magnitude of the scale.

Results: The final scale consists of 25 items using 3 stimuli (grandparent’s view of grandchild, mother of grandchild, father of grandchild) measuring a single domain (overall view of family). The Grandparent Perceptions of Family Scale (GPFS) includes 5 factors for a total explained variance of 60%. Coefficient α for the total scale is .90.

Discussion: The GPFS should be helpful to clinicians and researchers in their assessment of grandparents within multigenerational families and in planning appropriate interventions to address the roles and relationships of grandparents within families. Further GPFS evaluation is suggested with grandparents of diverse racial or ethnic identification; social economic position; education; and family definition, roles, composition, and structure.

Key Words: family · grandparents · psychometric testing

Grandparents play a potentially powerful role in the health of families (Cherlin & Furstenberg, 1986; Pillemer & White, 1989; Taylor, Chatters, & Jackson, 1993), and particularly in the health of their grandchildren (Boon & Brussoni, 1998; Hurme, 1997; Rempusheski, 1990). Reciprocal grandparent-grandchild relationships benefit grandparents as well, particularly as they age (Camp et al., 1997). The grandparent role and relationship with a grandchild are often viewed as important by middle-aged working men and women (Reitzes & Mutran, 2002) and as most positive in later life, at a time when older adults experience many role losses, such as retirement and spousal death (Rempusheski, 1990). The age of onset of grandparenthood is influenced by intergenerational population trends and interfamilial behavioral patterns and is estimated at 50 years in the United States (Szinovacz, 1998). New custodial grandparent questions in the 2000 U.S. Census will begin to capture trends and patterns of grandparents in multigenerational households (Simmons & Dye, 2003) and help determine prevalence, social demographic characteristics, and service use by custodial grandparents (Minkler & Fuller-Thomson, 2005).

Assessing grandparents’ perceptions of family is important for three reasons. First, the increase in the percentage of older people in industrialized countries is contributing to an increasing number of three- and four-generation families—more than ever before. Second, rarely are grandparents’ perceptions assessed, acknowledged, or viewed by healthcare professionals in the context of the potential effects on health, particularly in everyday stressful family events, such as birth of a child, divorce, illness, and death.

Veronica F. Rempusheski, PhD, RN, FAAN, is Associate Professor, The University of Kansas School of Nursing, Kansas City. Catherine T. O’Hara, MS, RN, ANP, is Adult Nurse Practitioner, Joslin Diabetes Center, State University of New York, Upstate Medical University, Syracuse, New York.
of a family member. Third, in healthcare and particularly for nurses at the front line of caring for individuals and families, it is becoming essential to take into consideration the dynamics, influences, and perceptions of members of each generation in everyday events and the effects of these situations on individual and family health.

Few researchers have addressed the broad and most prevalent role of grandparenting within three- and four-generation families (Mueller, Wilhelm, & Elder, 2002; Rempusheski, 1990; Werner, Lowenstein, & Katz, 1998). Given the diversity of definitions of family composition, understanding the majority role of grandparents within families is a challenge; the first step in achieving this understanding is developing a measure to capture the perception of family from the perspective of the grandparent. The purpose of this article is to describe the development and psychometric evaluation of the Grandparent Perceptions of Family Scale (GPFS).

Background

Prior to the 1970s, grandparents were rarely the central focus of either individual or family studies and were depicted as peripheral (Kahana & Kahana, 1971). Authors presented grandparenthood as a role and a social relationship framed in social role theory or exchange theory—a role having attitudinal, symbolic, behavioral, and affective aspects. Most often, the grandparent role has been studied using a narrow definition and rarely have all the complexities of the role been examined (Werner et al., 1998). The concept of family as viewed in healthcare in modern American society in the 1970s and 1980s included mother and child and sometimes father, but rarely grandparents (Blackburn & Lowen, 1986; Cronenwett, 1985). Grandparents were on the outside peering into the world of the nuclear family, wanting desperately to enter and participate (Rempusheski, 1990).

Custodial grandparents became highly visible in the 1980s (Fuller-Thomson, Minkler, & Driver, 1997) and continue to dominate the literature on grandparents (American Association of Retired Persons [AARP], 2002; Minkler & Fuller-Thomson, 2005). However, custodial grandparents are a majority of the population of grandparents. In 2002, 9% of African American children, 6% of Hispanic children, 4% of non-Hispanic White children, and 3% of Asian and Pacific Islander children were living in grandparent-headed households (Fields, 2003). Custodial grandparent roles and relationships within families are unique in situations of infants born with acquired immunodeficiency syndrome (AIDS), addictions, and neglected or abandoned by addicted, deceased, or incarcerated parents (Calandro & Hughes, 1998; Dressel & Barnhill, 1994; Minkler & Roe, 1993). Custodial grandparents have reported more compromised physical and mental health than noncustodial grandparents (Musil & Ahmad, 2002); however, health effects appear to vary by cultural norms and traditions associated with ethnic or racial differences (Goodman & Silverstein, 2002).

In the 1990s, intergenerational family theory influenced studies of grandparents (Werner et al., 1998). These studies recognize the influence of the grandparents’ meaning of grandparenthood on the other generations and the mediating relationship of the middle generation on both the first generation (grandparents) and the third generation (grandchildren; Holladay et al., 1997). An aggregate of each generation captures a full picture of interfamily relationships. The challenges to this theoretical perspective are the multiple configurations of families and how they might be represented in their relationships across and within generations. Factors influencing the nature of these relationships include age and health of grandparents and grandchildren (Mueller et al., 2002), geographic distance (Cherlin & Furstenberg, 1986), frequency of contact (Uhlenberg & Hammill, 1998), gender and lineage (Chan & Elder, 2000; Somary & Stricker, 1998), geographic region (Kivett, 1993), culture or race (Hurme, 1997; Minkler & Fuller-Thomson, 2003; Pruchno, 1999; Taylor et al., 1993), history of the relationship (Whitbeck, Hoyt, & Huck, 1993), and context of the family life cycle (Silverstein & Marenco, 2001).

Scale Development

Theoretical Grounding

The GPFS is theoretically grounded in an empirically derived model, Becoming a Grandparent. This model was generated from a grounded theory study of grandparents of premature newborns (Rempusheski, 1990). Grandparent perceptions of family emerged as a stressor factor, and were particularly relevant within the context of interfamily relationships of multigenerational families. Grandparent perceptions were conceptualized as having two domains. The first domain was an overall view, defined as a grandparent’s characterization of a member of the multigenerational family. This characterization may include a grandparent’s perception of the person’s physical and emotional characteristics, his or her personality, and contributions to society in general. The second domain was a relationship view, defined as a grandparent’s characterization of the relationships an individual family member exhibited within the multigenerational family. This characterization may include a grandparent’s perception of the nature of the relationship experienced or observed between this person and other members of the multigenerational family. In this model, grandparent perceptions of family and meanings of grandparenthood were hypothesized to have a relationship with characteristics of other generations. No scale existed to measure grandparent perceptions of family; therefore, it was essential to develop a scale to measure these, not only to test the model but also to help healthcare professionals assess multigenerational families in clinical arenas.

Scaling Method

The GPFS is a semantic differential scale. This type of scaling method was chosen because the shared cognition and knowledge of a cohort of people, such as grandparents, presumes a common core meaning to concepts. This scaling method (Osgood, Suci, & Tannenbaum, 1957) assumes a semantic space with multiple dimensions or factors that together account for the variance in meaningful judgments. The purpose of this scaling method is to differentiate in terms of scales. With appropriate analysis applied to the
scale or factor scores, conclusions can be drawn about similarities for a set of concepts for a group. The concepts are referred to as stimuli to be differentiated with the use of bipolar adjectives. Three dominant factors have appeared in the multiple analyses conducted on semantic differential scales. These factors are evaluative, potency, and action. Ideally, a tool should contain at least three scales for each factor to provide the subject with a balanced space (e.g., three evaluative, three potency, and three action). The evaluative factor accounts for the largest variance. The actual factor composition, although proposed in the development of a tool, is not known until after testing and factor analysis have been conducted (Osgood et al.). The theoretical, domain, and semantic rules applied to the GPFS must be kept in mind in the interpretation of the factor analysis. Overall, the semantic differential scaling method is evaluated as easy to use and versatile in the clinical area (Kerlinger, 1973; Nunnally & Bernstein, 1994; Osgood et al.). Participants can be expected to make judgments at a rate of 10–20 items per minute. However, the exact timing will vary with the age and intellectual ability of the participants and the types of concepts and scales being used (Osgood et al.). A semantic differential scale may consist of 5–9 scale steps; a 7-step scale is considered optimal (Waltz, Strickland, & Lenz, 1991).

**GPFS Structure**

There were 70 items (10 adjective pairs for each of seven stimuli), and two domains (overall view and relationship) in the original GPFS. The domains, concepts (stimuli), and descriptors underlying adjective pairs were derived inductively (Rempusheski, 1990). Adjective pairs were chosen for their factorial composition, relevance to the concepts, semantic stability (stable meaning across a set of concepts), and linearity between polar opposites. Linearity as used here means that from a semantic scaling theory perspective, a zero point or point of origin must be present for each adjective pair (Osgood et al., 1957). The items were arranged on a 7-step scale labeled extremely, quite, and slightly for each bipolar adjective in the adjective pair and a central neutral point labeled neither. The items included seven stimuli concepts representing members of a three-generation family, arranged within the two domains (the grandparent’s view of himself, my grandchild, my adult child, my adult child-in-law; the grandparent’s view of relationship with my grandchild, my adult child, my adult child-in-law), and 22 adjective pairs. Three adjective pairs were in the action subscale, 11 adjective pairs in the evaluative subscale, and 8 in the potency subscale.

**Face and Content Validity**

The 70-item GPFS was piloted with 12 community-dwelling, cognitively intact grandparents (8 women, 4 men) to determine its face validity (outward appearance), readability, variability, and relevance to the layperson. These grandparents represented a 30-year range in age (ages 50s to 80s), were grade school to college educated, and had 1–20 years experience as grandparents. Each was sent a copy of the GPFS and a letter of invitation requesting their anonymous responses about the form’s content and readability after completing the form. They were asked to make their comments on the basis of their experience as a grandparent and to return comments and the form in the enclosed self-addressed envelope. All 12 grandparents responded.

Content validity was assessed from a panel of six professionals with content expertise in intergenerational families and research experience in instrument development. These experts were asked to evaluate the GPFS for domain and item clarity, interpretation of terms, and item relevance on a 4-point scale of not relevant (scored 1) to very relevant (scored 4). Indices of content validity (CVI) were computed. CVI ranged from .60 to 1.0.

Suggestions for changes and additions were sought from the lay and expert panels. Adjective pairs, including additions and suggestions received from panels, were refined using card sort to distinguish bipolar terms and further refined for their semantic space and linearity. For example, the healthy–sick adjective pair was changed to healthy–unhealthy to decrease the semantic space in this adjective pair and to make its semantic space similar to other items in the scale. Irrelevant adjectives were deleted and suggested relevant adjectives were added. Stimuli were refined to align with language in the literature and to remove age and family structure biases. For example, the view of my adult child and view of my adult child-in-law stimuli were changed to view of mother of my grandchild and view of father of my grandchild. Family experts suggested adding the concept of couple to the relationship domain with the rationale that the grandparent’s perception of an individual parent may be different than the grandparent’s perception of both parents as a couple. The revised GPFS was organized into two domains and comprised 112 items within 8 stimuli and 27 adjective pairs.

### Table 1. GPFS Sample Items

<table>
<thead>
<tr>
<th>How I View My Grandchild</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td></td>
</tr>
<tr>
<td>Active</td>
<td></td>
</tr>
<tr>
<td>Emotionally strong</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** GPFS = The Grandparent Perceptions of Family Scale; E = evaluative; A = action; and P = potency.
The action, potency, and evaluation semantic subscales contained 3, 11, and 13 adjective pairs, respectively. Table 1 illustrates sample items—one each of an evaluative, action, and potency adjective pair.

Methods

Population and Sample

The population of interest for evaluating the GPFS were community-dwelling grandparents who had a minimum of one grandchild; were 35 years of age or older; and could read, write, and speak English. Following approval by the human protection committee of the university and signed informed consent from participants, data were collected from community-dwelling grandparents in two northeastern states. Grandparents with more than one grandchild were asked to select one grandchild for their focused responses on the GPFS and to write their rationale for the selection on the demographic form. The 551 grandparents who responded to written announcements in local papers, university newsletters, Web ads, group presentations (women’s or men’s social groups), and referrals were given the GPFS, a demographic form (GPDF) and a Health Status Questionnaire (HSQ; 1994). A total of 310 forms were returned—a 56% response rate, of which 306 GPFS and GPDF forms and 305 HSQs were usable in the final analysis. A random sample of approximately 10% of the total sample was selected to repeat the GPFS 1 month following the initial completion. Thirty-three GPFS forms were sent as retests. A total of 27 GPs returned the retest—an 82% response rate, of which 25 forms were usable for stability evaluation.

Instruments

GPFS. The 112-item GPFS, as previously described, was formatted in size 14 font print and was estimated to take 20 min to complete, taking into consideration estimates from previous analyses (Osgood et al., 1957) and eye and central nervous system changes associated with aging (Kline, Schieber, Abusamra, & Coyne, 1983). Items were scored 1 (for the least desirable attribute in the adjective pair) to 7 (for most desirable characteristic) on the 7-point scale. Items were summed for a total score. The higher the score, the more favorable is the grandparent's perception of the relationship and view of the family.

GPDF. The 18-item GPDF was constructed for this study and was based on grandparent, family, and selected grandchild demographic factors cited in the literature (Cherlin & Furstenberg, 1986; Rempusheski, 1990) as being most relevant in intergenerational family relations. Data were indexed at the nominal (e.g., gender, ethnicity, religion), ordinal (e.g., education, economic status), and interval (e.g., age, number of family members in each generation, geographic distance between grandparent and selected grandchild) levels. An open-ended question requested each participants to write his or her reason for choosing the grandchild.

HSQ. Grandparents’ health is often cited as influencing interfamilial relationships (Mueller et al., 2002) wherein family members may perceive grandparents with chronic disease or elderly grandparents as unable (physically/ functionally or emotionally) or uninterested, due to their limitations, in family interactions. For this reason, the HSQ version 2.0 (1994) was used to measure grandparents’ perceptions of their health, as an added characteristic to define the sample. The HSQ has 39 forced-choice questions assessing eight health attributes: health perception, physical functioning, role limitations attributed to physical health (role-physical), role limitations attributed to emotional problems (role-emotional), social functioning, mental health, bodily pain, and energy/fatigue. In addition, the HSQ includes an index of health status change and a screening for depression. The first 36 items of the HSQ are comparable to the RAND 36-item Health Survey 1.0 (RAND Health Sciences Program, 1992) and the MOS 36-item short form (SF-36; Ware, 1993). All three tools, with the same 36 items, have been used widely to measure health status of adults and older adults in the general population and with patients with various chronic conditions. Internal consistency reliability coefficients have ranged from .81 to .86 on each of the SF-36 subscales (Stewart et al., 1989). Radosevich, Wetzler, and Wilson (1994) reported Cronbach’s coefficient on the HSQ subscales that ranged from .75 to .89 for a sample of 2,384 respondents aged 18–64 years and .79 to .93 for a sample of 1,814 respondents aged 65 and older. Scoring HSQ items entails recoding responses to weighted values equating 100 points per item, summing across items, and calculating averages for each of the eight health attribute scales. The three-item depression screener in the HSQ is clinically valid and reliable for identifying at-risk (high, low) for depression and melancholia (Rost, Burnam, & Smith, 1993). Rost et al. reported 81% sensitivity, 95% specificity, 33% positive predictive value, and 99% negative predictive value using the three-item depression screener in samples of patients with medical and mental health diagnoses. Completion of the HSQ in size 14 font print was estimated to take 10 min.

Psychometric Analysis Procedures

Descriptive analyses were performed on the GPDF, HSQ, and GPFS data. Frequency distribution of items was assessed. Coefficient alpha reliabilities were computed for HSQ subscales. A principal components factor analysis with varimax rotation was performed to evaluate construct validity, structure, and magnitude of the GPFS. The goal was instrument parsimony with structure of the scale and magnitude of items determined through the factor analysis. A cutoff point of .50 was used for factor loading. Each factor must account for at least 5% of the variance and have an eigenvalue of ≥1. Reliability procedures were conducted for internal consistency of the GPFS and stability. Evaluation criteria for internal consistency were established for 5 sets of correlations: (a) item–item (.30–.70), (b) item–subscale (.50–.65), (c) item–total scale (.30), (d) subscale–subscale (.40–.65), and (e) subscale–total scale (.55), and the coefficient alpha for the scale and subscales (.70; Nunnally & Bernstein, 1994). Stability was evaluated with test–retest coefficient (.70) for items and scales (N = 26) with a paired t test (p < .50, 95% confidence interval) to evaluate the difference between Time 1 and Time 2 responses on the GPFS.
Results

Sample
Grandparent participants ranged in age from 41 to 92 years (M = 64, SD = 9.8); 72% were female, 44% were retired, 62% had a minimum of an undergraduate degree from a college or university, 55% rated their economic status as comfortable, and 71% were living with a spouse or partner. Participants reported that the age they became a grandparent ranged from 26 to 74 years (M = 52, SD = 7.5); they had 1Y9 children (M = 3, SD = 1.4) and 1Y18 grandchildren (M = 4, SD = 3.1). Most (93%) were born in the United States yet identified themselves with a broad set of descriptors including other countries, skin color, and religion. Seven percent of the sample immigrated to the United States from countries within Europe (Germany, Ireland, Italy, Poland, United Kingdom), the Americas (Canada, Central America), and the Caribbean (Jamaica, Puerto Rico). Aggregated self-reported ethnic identities resulted in a sample composition of 97% White, 2% Black, and 1% Hispanic. Ninety-one percent declared a religious affiliation, of which 46% stated Protestant denominations, 39% Catholic, and 15% Jewish.

HSQ health dimension scale scores and alpha reliabilities are presented in Table 2. In general, age-associated declines in health were greater for the scales that measure physical health (i.e., physical functioning, role-physical) than mental health (i.e., role-emotional, mental health), were not age-associated for health perception, and were consistent with nationally aggregated data on middle age and older adult samples (Radosevich et al., 1994). Participants rated their health the same compared to 1 year ago (M = 2.86, SD = 0.638) and most of the sample was at low risk for depression (82%), melancholia (83.3%), or a major depression superimposed on dysthymia (91%).

Selected grandchildren were aged 0.1Y37 years (M = 6.5, SD = 7.0), equally male (50.2%) and female (49.8%), and lived 0 (same household or same street) to 3,000 miles from the grandparent participants (M = 284.4, SD = 657.9). Grandparents stated that reasons for selecting a grandchild were because he or she was their oldest/first (39.5%), only (19.4%), or youngest grandchild (7.3%); had the most contact with, lived with, or were geographically closest to the grandparent (13.8%); or other reasons that described favored grandchild characteristics (20.0%). These written responses revealed that approximately 3% of the sample were part-time caregivers for a grandchild. For example, two grandparents noted death at childbirth of the mother of a grandchild whom they cared for with the grandchild’s father, and seven others wrote that they babysat daily to several times per week. Nearly all of the grandparents responded to the mother of my grandchild and father of my grandchild stimuli and described the parents of the selected grandchild as widowed, married, living together, separated, or divorced. A few grandparents did not complete the view of father of my grandchild items and responded that they had no relationship with the father of their grandchild because of the father’s absence or

<table>
<thead>
<tr>
<th>Health Dimension Scale</th>
<th>α</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health perception</td>
<td>.723</td>
<td>74.88</td>
<td>16.31</td>
</tr>
<tr>
<td>Physical functioning</td>
<td>.887</td>
<td>81.46</td>
<td>19.76</td>
</tr>
<tr>
<td>Role-physical</td>
<td>.883</td>
<td>80.39</td>
<td>33.74</td>
</tr>
<tr>
<td>Role-emotional</td>
<td>.815</td>
<td>86.77</td>
<td>28.73</td>
</tr>
<tr>
<td>Social functioning</td>
<td>.805</td>
<td>90.26</td>
<td>17.56</td>
</tr>
<tr>
<td>Mental health</td>
<td>.802</td>
<td>80.60</td>
<td>13.98</td>
</tr>
<tr>
<td>Bodily pain</td>
<td>.835</td>
<td>81.16</td>
<td>20.00</td>
</tr>
<tr>
<td>Energy/fatigue</td>
<td>.835</td>
<td>67.16</td>
<td>16.57</td>
</tr>
</tbody>
</table>

Note. HSQ = Health Status Questionnaire.

TABLE 3. GPFS Factor Analysis

<table>
<thead>
<tr>
<th>Stimulus: “View of”</th>
<th>Number of Items</th>
<th>Number of Items ≥ .50</th>
<th>Number of Factors</th>
<th>Total % of the Variance</th>
<th>Number of Items Deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myself</td>
<td>18</td>
<td>18</td>
<td>6</td>
<td>59.3</td>
<td>0</td>
</tr>
<tr>
<td>GC</td>
<td>18</td>
<td>17</td>
<td>5</td>
<td>62.9</td>
<td>1</td>
</tr>
<tr>
<td>Mother of GC</td>
<td>18</td>
<td>16</td>
<td>4</td>
<td>58.5</td>
<td>2</td>
</tr>
<tr>
<td>Father of GC</td>
<td>18</td>
<td>16</td>
<td>4</td>
<td>60.0</td>
<td>2</td>
</tr>
<tr>
<td>Relationship with GC</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>65.0</td>
<td>0</td>
</tr>
<tr>
<td>Relationship with mother of GC</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>72.3</td>
<td>0</td>
</tr>
<tr>
<td>Relationship with father of GC</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>73.6</td>
<td>0</td>
</tr>
<tr>
<td>Relationship with mother and father of GC as a couple</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>69.6</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. GPFS = The Grandparent Perceptions of Family Scale; GC = grandchild.
unknown location in situations of divorce, separation, or abandonment.

Construct Validity
Consistent with semantic differential approaches for analysis, each grouping of stimulus-to-set of adjective pairs was evaluated for factors. Table 3 summarizes the results of the factor analysis. In the first 4 groupings, 5 items were deleted. In the relationship groupings, all items loaded on one factor for each and met the criteria; so initially none of these items was deleted.

Internal Consistency and Stability
When evaluated for internal consistency, two stimulus-adjective sets containing 18 items each (view of myself, view of mother and father as couple) fell out. The myself stimulus-adjective items did not meet minimum criterion, indicating that set’s irrelevance to the domain. However, view of mother and father as couple stimulus-adjective items exceeded criterion, indicating its redundancy to mother of my grandchild and father of my grandchild stimuli-adjective sets.

The deletion of items overlapped with each of the procedures. All items and their failure to meet a stipulated criterion (e.g., item–item, item–subscale, item to total scale) were arranged in a matrix for comparison purposes. For example, some items did not meet one of these criteria whereas other items may not have met two or three criteria.

Regarding GPFS stability, the test–retest coefficient for 69 items was under .70 and considered weak. Forty-three items ranged from .70 to .96, and the subscales ranged from .73 to .96. In t test evaluation (confidence interval of 95%), 3 items had significant differences, indicating instability whereas all subscales were stable between Time 1 and Time 2.

When all items in the matrix were evaluated, the revised GPFS consisted of 26 items (N = 272 cases) within a single domain, overall view of family member, with 3 stimuli: grandchild, mother of my grandchild, father of my grandchild, and 13 adjective pairs. High correlations between relationship items and overall view items signaled redundancy of domains and resulted in aggregating the strong remaining items into one domain of overall view of family member.

A principal components factor analysis with varimax rotation of this revised scale resulted in 25 items loading on 5 factors at ≥.50, with an eigenvalue ≥1, minimum of 5% variance per factor, for a total explained variance of 60%. The coefficient for the total scale was .90 and did not change with one item deleted. Table 4 summarizes the factor loadings, which are rated good (.52) to excellent (>70; Nunnally & Bernstein, 1994). One item (#12) loaded on two factors (Factor 3 at .52 and Factor 5 at .52). This item fits conceptually with Factor 3. Item 11 did not meet the criterion of ≥.50 and was deleted.

One adjective pair was problematic because of its semantic space (trustworthy/suspicious) and its conceptualization according to comments made by subjects, so in the final version of the GPFS, this item has a smaller semantic space, consistent with the other GPFS adjective pairs, and is stated as trusting/nontrusting. The final scale is one domain and 25 items, with each item having a stimulus and adjective pair (Table 5). Coefficient for the total scale was .90.

Discussion
Measures such as the GPFS are essential so as to further knowledge and understanding about multigenerational families, factors that may influence an individual family member’s health, and factors that may influence the health of a family. When compared to U.S. Census data, the study sample was more educated, more economically comfortable, and less racially and ethnically diverse than same age persons in the United States. Sixty-two percent of the study sample, with a mean age of 64 years, had earned a minimum of a baccalaureate degree compared to 20.31% nationally for persons aged 60–64 years, and 27.4% and 33.2% for the two northeastern states within which the participants resided (U.S. Census Bureau, 2003). All participants spoke English. A majority of subjects

| TABLE 4. GPFS-25 Factor Number and Name, (Item Number), and Factor Loadings per Item |
|--------------------------------|--------------------------------|--------------------------------|----------------------------|----------------------------|
| Factor 1 Father of GC Evaluation | Factor 2 GC Evaluation | Factor 3 Mother of GC Evaluation | Factor 4 GC Potency | Factor 5 Mother of GC Potency |
| (23).77 | (8).63 | (16).74 | (6).69 | |
| (24).56 | (10).63 | (17).77 | (9).64 | |
| (25).64 | | (19).75 | | |
| (26).76 | | | | |

Note. GPFS = The Grandparent Perceptions of Family Scale; GC = grandchild.
were born in the United States, of European heritage, and White.

The northeast United States has the highest educational attainment of all four regions of the nation, has the least percentage of Hispanic residents, and is second highest only to the Western region ($46,800) in median household income ($46,700; Weinberg, 2004). Consequently, the economic status of the study sample is neither reflective of a large portion of the population that resides in the Midwest and Southern regions of the nation nor of residents in the Northeast who are below the median income. The GPFS should be evaluated with grandparents who are minimally educated at elementary and high school levels, are economically disadvantaged, and represent racial or ethnic populations, particularly in the Midwest and Southern United States.

The GPFS should be translated into Spanish and psychometrically evaluated with the fastest growing racial or ethnic group, Hispanic, that is projected to represent 15.5% of the total U.S. population by 2010, compared to a projected 13.1% African American and 4.6% Asian (U.S. Census Bureau, 2004, June 14). The health status of middle aged and elderly persons in this sample illustrated a range from high functional and physical health to lower levels of function, predominately along age lines consistent with national samples (Radosevich et al., 1994). Subjects self-selected to participate in this study and may represent those grandparents who have experienced a lifespan adaptation in their grandparent–grandchild relationship despite health or functional changes (Silverstein & Marenco, 2001).

GPFS Factor 1 (Father of Grandchild Evaluation) and Factor 3 (Mother of Grandchild Evaluation) have five adjective pairs in common; however, two adjective pairs are unique to Factor 1 (warm–cold, responsible–irresponsible) and one adjective pair is unique to Factor 3 (easy–difficult). The most enduring characteristic of fatherhood is providing economically for a family and secondarily as contributing an emotional or compassionate attribute, illustrating a growing but small proportion of men who are significantly involved in nurturing their children (Dowd, 2000).

### TABLE 5. GPFS-25: Stimulus and Adjective Pair Items for Each Factor

<table>
<thead>
<tr>
<th>Factor Number</th>
<th>Item Number</th>
<th>Stimulus</th>
<th>Adjective Pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>How I view father of my GC</td>
<td>Thoughtful–thoughtless</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td></td>
<td>Loving–unloving</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td></td>
<td>Affectionate–distant</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td></td>
<td>Warm–cold</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td></td>
<td>Responsible–irresponsible</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td></td>
<td>Trusting–nontrusting</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td></td>
<td>Kind–cruel</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>How I view my GC</td>
<td>Affectionate–distant</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>Loving–unloving</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
<td>Happy–sad</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td>Flexible–rigid</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>How I view mother of my GC</td>
<td>Trusting–nontrusting</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td></td>
<td>Easy–difficult</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td></td>
<td>Kind–cruel</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td></td>
<td>Loving–unloving</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td></td>
<td>Affectionate–distant</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td></td>
<td>Thoughtful–thoughtless</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>How I view my GC</td>
<td>Easy–difficult</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>Thoughtful–thoughtless</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
<td>Nonmanipulative–manipulative</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>Trusting–nontrusting</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td></td>
<td>Emotionally strong–emotionally weak</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>How I view mother of my GC</td>
<td>Emotionally strong–emotionally weak</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td></td>
<td>Happy–sad</td>
</tr>
</tbody>
</table>

*Note.* GPFS = The Grandparent Perceptions of Family Scale; GC = grandchild. Items in Factors 1–3 are Evaluative; items in Factors 4–5 are Potency. No Action subscales emerged from the factor analysis.
distinguishing adjective pairs in Factor 1 appear to be addressing the compassionate (warm–cold) and economic responsibility (responsible–irresponsible) characteristics that dominate current societal expectations of fathers. Also, these adjective pairs may be reflective of the good dad–bad dad complex that describes the contrasting sides of contemporary fatherhood (Pleck, 2004). The good dad is both provider and partner in childcare, whereas the bad dad denies paternity or refuses to support his children.

The adjective pair that distinguishes Factor 3 (easy–difficult) may reflect the grandparents’ perception of the middle generation, specifically mothers of grandchildren, as mediators of the grandparent–grandchild relationship and gatekeepers of access to grandchildren (Holladay et al., 1997; Mueller et al., 2002; Silverstein & Marenco, 2001; Whitbeck et al., 1993). Maternal family lines tend to be strongest in grandparent–grandchild relationships and 72% of the study sample was women; therefore, this distinguishing adjective pair may characterize a matrilineal advantage (Chan & Elder, 2000; Mueller et al., 2002) in each grandparent’s view of his or her relationship with the mother of the grandchild. Grandparents may view their relationship with the mother of a grandchild as difficult if access to a grandchild is restricted or curtailed or as easy if their access to a grandchild is facilitated or invited.

The items of the GPFS were inductively derived and supported by the Becoming a Grandparent theoretical model (Rempusheski, 1990). The tool appeared relevant to grandparents who responded to the gender-specific parental stimuli and completed the GPFS. Their families comprised grandchildren of married, divorced, separated, and widowed men and women parents. On the basis of its psychometric evaluation, the GPFS is a reliable, valid, and parsimonious self-administered tool that can be used to quickly assess grandparents’ perceptions of families with similar structures and roles as those of this sample.

Further testing of the GPFS with gender-neutral parental stimuli is suggested to measure grandparent perceptions of grandchildren and parents of grandchildren within families with gay and lesbian partners, single heterosexual men and women with adopted children, and single heterosexual and homosexual women who conceive children via donor insemination. Fulcher, Chan, Raboy, and Patterson (2002) reported that the frequency of contact between grandchildren and grandparents was similar for children of lesbian mothers and for children of heterosexual parents for both biological and nonbiological grandparents; however, they did not address grandparent perceptions of family. Future development of family perception scales need to take into account these family structures and the theoretical, conceptual, and measurement issues associated with capturing the multiple definitions of parents and family.

Additional validity testing of the GPFS is suggested, such as convergent–discriminant validity with a family functioning scale. This evaluation is of particular importance as the scale is translated into other languages and revised to account for diversity of educational attainment, socioeconomic position, ethnic or racial composition, and family structure.  

References


Exercise
Effectiveness of a behavioral change intervention in Thai elders after knee replacement (Harnirattisai, Johnson), 97
Midlife women’s adherence to home-based walking during maintenance (Wilbur, Vassallo, Chandler, McDevitt, Miller), 33

Factor Analysis
Creation of a model of independence for community-dwelling elders in the United States (Baker), 288

Family
Evaluating in-home training for fathers of children with autism using single-subject experimentation and group analysis methods (Elder, Valcente, Yarandi, White, Elder), 22
Psychometric properties of the Grandparent Perceptions of Family scale (Rempe-Sheski, O’Hara), 419

Family Intervention
Family education and support interventions in heart failure: a pilot study (Dunbar, Clark, Deaton, Smith, De, O’Brien), 158

Fatigue
Fatigue and fatigue-relieving strategies used by Hong Kong Chinese patients after hematopoietic stem cell transplantation (So, Tai), 48

Fetus
Effects of immediate versus delayed pushing during second-stage labor on fetal well-being: a randomized clinical trial (Simpon, James), 149

Gender
Anger dysregulation, depressive symptoms, and health in married women and men (Carrèr, Mittmann, Woodin, Tabares, Yoshimoto), 184

Geriatric Population
Clinical outcomes of aging in place (Marek, Popejoy, Petroski, Mehr, Rantz, Lin), 202
Creation of a model of independence for community-dwelling elders in the United States (Baker), 288
Developing and testing a self-care intervention for older adults in caregiving roles (Teel, Leenerts), 193

Effectiveness of a behavioral change intervention in Thai elders after knee replacement (Harnirattisai, Johnson), 97
A framework for studying the nutritional health of community-dwelling elders (Chen), 13
Older widows’ experience of home care (Porter), 296
Reducing assaults against nursing home caregivers (Gates, Fitzwater, Succop), 119

Grandparents
Psychometric properties of the Grandparent Perceptions of Family scale (Rempe-Sheski, O’Hara), 419

Guest Editorial (Department)
American Nurses Foundation: celebrating 50 years of service 1955-2005 (ANF Board of Trustees), 147
Revising manuscripts and grants (Dow), 217

Health and Retirement Study
Profiles of self-rated health in midlife adults with chronic illnesses (Finnegan, Marion, Cox), 167
Healthcare providers
Evaluating the psychometric competence assessment instrument among healthcare providers (Doorenbos, Verbitsky, Given, Given), 229

Health Disparities
Health disparities, social injustice, and the culture of nursing (Giddings), 304
Testing the theory of planned behavior to predict mammography intention (Steele, Porche), 332

Health Interventions
Culturally tailored diabetes education program for Chinese Americans: a pilot study (Wang, Chan), 347

Use of time studies for determining intervention costs (Findorff, Wyman, Croghan, Nyman), 280

Health Promotion
Developing and testing a self-care intervention for older adults in caregiving roles (Teel, Leenerts), 193

Health-related Quality of Life
Testing the psychometric properties of the Minnesota Living With Heart Failure Questionnaire (Heo, Moser, Riegel, Hall, Christman), 265

Health Status
Anger dysregulation, depressive symptoms, and health in married women and men (Carrèr, Mittmann, Woodin, Tabares, Yoshimoto), 184
Creation of a model of independence for community-dwelling elders in the United States (Baker), 288
Health disparities, social injustice, and the culture of nursing (Giddings), 304
Heart Failure
Family education and support interventions in heart failure: a pilot study (Dunbar, Clark, Deaton, Smith, De, O’Brien), 158
Testing the psychometric properties of the Minnesota Living With Heart Failure Questionnaire (Heo, Moser, Riegel, Hall, Christman), 265

Hemopoietic Stem Cell Transplantation
Fatigue and fatigue-relieving strategies used by Hong Kong Chinese patients after hematopoietic stem cell transplantation (So, Tai), 48

Hemostasis
Effects of three groin compression methods on patient discomfort, distress and vascular complications following a percutaneous coronary intervention procedure (Chdan, Sabo, Savik), 391

H&H Lactation Scale
A psychometric assessment of the H&H Lactation Scale among a sample of Thai mothers using a repeated measurement design (Punthmatharith, Singh), 313
Hierarchical Linear Models
An analytic strategy for modeling multiple-item responses: a breast cancer symptom example (Doorenbos, Verbitsky, Given, Given), 229
Impact of centering first-level predictors on individual and contextual effects in multilevel data analysis (Wu, Woodruff), 212

Hierarchical Models
The impact of hospital nursing characteristics on 30-day mortality (Easterbrook, Miodoz, Cummings, Rickert, Giovannetti), 74

Hierarchical Structure
Multilevel modeling of a clustered continuous outcome: nurses’ work hours and burnout (Park, Lake), 406

HIV/AIDS
A randomized clinical trial of an HIV-risk-reduction intervention among low-income Latina women (Peragallo, DeForge, O’Campo, Lee, Kim, Cianelli, Ferrer), 108

Home Care Services
Older widows’ experience of home care (Porter), 296

Hospital(s)
A meta-analysis of the effect of a hospital-based case management on hospital length-of-stay and readmission (Kim, Soeken), 255

Hospital Restructuring
Mitigating the impact of hospital restructuring on nurses: the responsibility of emotionally intelligent leadership (Cummings, Hayduk, Estabrooks), 2

Independence
Creation of a model of independence for community-dwelling elders in the United States (Baker), 288

Injury
Dating violence in college women: associated physical injury, healthcare usage, and mental health symptoms (Amir, Gennaro), 235
Predicting nosocomial bloodstream infections using surrogate markers of injury severity (El-Masri, Hammad, Fox-Wawryshyn), 273

In Memoriam
Downs, Florence, 373

Institutional Ethics
Research environments that promote integrity (Jeffers), 63

Instrument Development
Evaluating content validity for children’s self-report instruments using children as content experts (Stewart, Lynn, Mishel), 414

Insulin Resistance
African American women, metabolic syndrome, and a National Cholesterol Education Program criteria: a pilot study (Appel, Floyd, Giger, Weaver, Luo, Hannah, Oraville), 339

Integrative Review
Combining evidence in nursing research: methods and implications (Whittenmore), 56

Interaction Model of Client Health Behavior
Profiles of self-rated health in midlife adults with chronic illnesses (Finnegan, Marion, Cox), 167

Item Response Theory
An analytic strategy for modeling multiple-item responses: a breast cancer symptom example (Doorenbos, Verbitsky, Given, Given), 229

Knee Replacement
Effectiveness of a behavioral change intervention in Thai elders after knee replacement (Harnirattisai, Johnson), 97

Labor and Delivery
Effects of immediate versus delayed pushing during second-stage labor on fetal well-being: a randomized clinical trial (Simpon, James), 149

Lactation
A psychometric assessment of the H&H Lactation Scale in a sample of Thai mothers using a repeated measurement design (Punthmatharith, Singh), 313

Latin Population
A randomized clinical trial of an HIV-risk-reduction intervention among low-income Latina women (Peragallo, DeForge, O’Campo, Lee, Kim, Cianelli, Ferrer), 108

Leadership
Mitigating the impact of hospital restructuring on nurses: the responsibility of emotionally intelligent leadership (Cummings, Hayduk, Estabrooks), 2

Length of Stay
A meta-analysis of the effect of a hospital-based case management on hospital length-of-stay and readmission (Kim, Soeken), 255

Letters to the Editor (Department)
Response to Mar Blegen’s editorial in the March/April issue of Nursing Research (Dreher), 218

Life History Study
Health disparities, social injustice, and the culture of nursing (Giddings), 304

Logistic Regression
Logistic regression and Bayesian networks to study outcomes using large data sets (Lee, Abbott, Johantgen), 133

Long-term Care
Clinical outcomes of aging in place (Marek, Popejoy, Petroski, Mehr, Rantz, Lin), 202
Reducing assaults against nursing home caregivers (Gates, Fitzwater, Succop), 119

Magnet Hospitals
The magnet process and the perceived work environment of nurses (Cimiotti, Quinlan, Larson, Pastor, Lin, Stone), 384
Scientific Misconduct
Research environments that promote integrity (Jeffers), 63
Self-care
Developing and testing a self-care intervention for older adults in caregiving roles (Teel, Leenerts), 193
Self-efficacy
Predictors of professional nursing practice behaviors in hospital setting (Manojlović), 41
Self-rated Health
Profiles of self-rated health in midlife adults with chronic illnesses (Finnegan, Marion, Cox), 167
Self-reports
Evaluating quality of life in children with cancer using children’s self-reports and parent-proxy reports (Yeh, Chang, Chang), 354
Shared Governance
The staff nurse Decisional Involvement Scale: report of psychometric assessments (Havens, Vasey), 376
Single-Subject Experimentation
Evaluating in-home training for fathers of children with autism using single-subject experimentation and group analysis methods (Elder, Valcante, Yarandi, White, Elder), 22
Smoking Cessation
Effectiveness of a nursing inpatient smoking cessation program in individuals with cardiovascular disease (Chounard, Robichaud-Ekstrand), 243
Social Ability
Creation of a model of independence for community-dwelling elders in the United States (Baker), 288
Social Justice
Health disparities, social injustice, and the culture of nursing (Giddings), 304
Sodium Restriction
Family education and support interventions in heart failure: a pilot study (Dunbar, Clark, Deaton, Smith, De, O’Brien), 158
Staffing
The impact of hospital nursing characteristics on 30-day mortality (Estabrooks, Midodzi, Cummings, Ricker, Giovannetti), 74
Structural Empowerment
Predictors of professional nursing practice behaviors in hospital setting (Manojlović), 41
Structural Equation Model
Mitigating the impact of hospital restructuring on nurses: the responsibility of emotionally intelligent leadership (Cummings, Hayduk, Estabrooks), 2
Surrogate Markers
Structural model for quality of life of patients with chronic cardiovascular disease in Korea (Han, Park, Park, Cheol), 85
Survey
Does it pay to pay? A randomized trial of pre-paid financial incentives and lottery incentives in surveys of nonphysician healthcare professionals (Ulrich, Danis, Kozol, et al), 178
Systematic Review
Combining evidence in nursing research: methods and implications (Whittemore), 56
Theory of Planned Behavior
Testing the theory of planned behavior to predict mammography intention (Steele, Porche), 332
Time Studies
Use of time studies for determining intervention costs (Fendt, Wyman, Croghan, Nyman), 320
Transplantation
Fatigue and fatigue-relieving strategies used by Hong Kong Chinese patients after hematopoietic stem cell transplantation (So, Tai), 48
Transtheroretical Model
Effectiveness of a nursing inpatient smoking cessation program in individuals with cardiovascular disease (Chounard, Robichaud-Ekstrand), 243
Treatment Fidelity
Treatment fidelity in behavior change research: a case example (Resnick, Inguito, Orwig, Yahiro, Hawkes, Werner, Zimmerman, Magaziner), 139
Treatment Implementation
Developing and testing a self-care intervention for older adults in caregiving roles (Teel, Leenerts), 193
Vascular Complications
Effects of three groin compression methods on patient discomfort, distress and vascular complications following a percutaneous coronary intervention procedure (Cihan, Sabo, Savik), 391
Violence
Dating violence in college women: associated physical injury, healthcare usage, and mental health symptoms (Amar, Gennaro), 235
Reducing assaults against nursing home caregivers (Gates, Fitzwater, Suocco), 119
Walking
Midlife women’s adherence to home-based walking during maintenance (Wilbur, Vassalo, Chandler, McDewitt, Miller), 33
Widowhood
Older widows’ experience of home care (Porter), 296
Women’s Health
African American women, metabolic syndrome, and National Cholesterol Education Program criteria: a pilot study (Appel, Floyd, Giger, Weaver, Luo, Hannah, Ovalle), 339
An analytic strategy for modeling multiple-item responses: a breast cancer symptom example (Doorenbos, Verbitsky, Given, Given), 229
Dating violence in college women: associated physical injury, healthcare usage, and mental health symptoms (Amar, Gennaro), 235
Effects of immediate versus delayed pushed delayed during second-stage labor on fetal well-being: a randomized clinical trial (Simpson, James), 149
Midlife women’s adherence to home-based walking during maintenance (Wilbur, Vassalo, Chandler, McDewitt, Miller), 33
A psychometric assessment of the H&H Lactation scale in a sample of Thai mothers using a repeated measurement design (Punthmarathirat, Singh), 313
A randomized clinical trial of an HIV-risk-reduction intervention among low-income Latina women (Peragallo, DeFoe, O’Campo, Lee, Kim, Canell, Ferrer), 108
Testing the theory of planned behavior to predict mammography intention (Steele, Porche), 332
Workplace Environment
The impact of hospital nursing characteristics on 30-day mortality (Estabrooks, Midodzi, Cummings, Ricker, Giovannetti), 74
The magnet process and the perceived work environment of nurses (Cimocotti, Quinlan, Larson, Paston, Lin, Stone), 384
Predictors of professional nursing practice behaviors in hospital setting (Manojlovic), 41
Reducing assaults against nursing home caregivers (Gates, Fitzwater, Suocco), 119

Author Index

Abbott, Patricia J.lauth See Lee, Sun-Mi
Amar, Angela F.; Gennaro, Susan; Dating Violence in College Women: Associated Physical Injury, Healthcare Usage, and Mental Health Symptoms, 235
American Nurses Foundation; American Nurses Foundation: Celebrating 50 Years of Service 1955–2005
Appel, Susan J.; Floyd, Natalie A.; Giger, Joyce N.; Weaver, Michael J.; Luo, Huabin; Hannah, Tiffany; Ovalle, Fernando; African American Women, Metabolic Syndrome, and National Cholesterol Education Program Criteria: A Pilot Study, 339
Baker, Margaret W.; Creation of a Model of Independence for Community-Dwelling Elders in the United States, 288
Bennett, Jill A.; The Consolidated Standards of Reporting Trials (CONSORT): Guidelines for Reporting Randomized Trials, 128
Blegen, Mary A.; Research Threat Changing at Risk, 1
Borse, Nagesh N.; See Doorenbos, Ardith Z.
Buetner, Linda J.ath See Kolanowski, Ann M
Carrère, Sybil; Mittmann, Angel; Woodin, Erica; Tabares, Amber; Yoshimoto, Dan; Anger Dysregulation, Depressive Symptoms, and Health in Married Women and Men, 184
Chan, Siu MA J.ath See Wang, Chen-Yen
Chandler, Peggy J.ath See Wilbur, Jollie
Chang, Chi-Weng J.ath See Yeh, Chao-Hsiung
Chang, Pi-Chen J.ath See Yeh, Chao-Hsiung

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
<table>
<thead>
<tr>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen, Cheryl Chia-Hui</td>
<td>A Framework for Studying the Nutritional Health of Older Adults With Cognitive Impairment, 13</td>
</tr>
<tr>
<td>Cheol, Kang Hyun</td>
<td>jt auth See Han, Kuen Sun</td>
</tr>
<tr>
<td>Chlan, Linda L; Sabo, Julie; Savik, Kay</td>
<td>Effects of Three Groin Compression Methods on Patient Discomfort, Distress, and Vascular Complications Following a Percutaneous Coronary Intervention Procedure, 391</td>
</tr>
<tr>
<td>Chouinard, Maud Christine; Robichaud-Ekstrand, Sylvie</td>
<td>The Effectiveness of a Nursing Inpatient Smoking Cessation Program in Individuals With Cardiovascular Disease, 243</td>
</tr>
<tr>
<td>Christman, Norma</td>
<td>jt auth See Hoe, Seongkum; Cimiotti, Jeannie P; Quinlan, Patricia M; Larson, Elaine L; Pastor, Diane K; Lin, Susan X; Stone, Patricia W; The Magnet Process and the Perceived Work Environment of Nurses, 384</td>
</tr>
<tr>
<td>Clark, Patricia C</td>
<td>jt auth See Dunbar, Sandra B; Cox, Cheryl</td>
</tr>
<tr>
<td>Crogan, Catherine F</td>
<td>jt auth See El-Mastri, Maher M</td>
</tr>
<tr>
<td>Cummings, Greta G</td>
<td>jt auth See Estabrooks, Carol A</td>
</tr>
<tr>
<td>Cummings, Greta; Hayduk, Leslie; Estabrooks, Carole</td>
<td>Mitigating the Impact of Hospital Restructuring on Nurses: The Responsibility of Emotionally Intelligent Leadership, 2</td>
</tr>
<tr>
<td>Danis, Marion</td>
<td>jt auth See Ulrich, Connie M; De, Anindya K</td>
</tr>
<tr>
<td>Dougherty, Molly C; Read Fiction to Know Nursing, 73</td>
<td></td>
</tr>
<tr>
<td>Dow, Karen H</td>
<td>jt auth See McNees, Patrick; Dow, Karen H</td>
</tr>
<tr>
<td>Dreher, Melanie C</td>
<td>Response to Mery Blegen’s Editorial in the March/April Issue of Nursing Research, 218</td>
</tr>
<tr>
<td>Dunbar, Sandra B; Clark, Patricia C; Deaton, Christi; Smith, Andrew L; De, Anindya K; O’Brien, Marian C</td>
<td>Family Education and Support Interventions in Heart Failure: A Pilot Study, 158</td>
</tr>
<tr>
<td>Elder, Jennifer H; Valacente, Gregory; Yarandi, Hosein</td>
<td>White, Deborah; Elder, Timothy H</td>
</tr>
<tr>
<td>Elder, Timothy H</td>
<td>jt auth See Elder, Jennifer H; El-Mastri, Maher M; Hammad, Tarek A; Fox-Wastlyn, Susan M; Predicting Nosocomial Bloodstream Infections Using Surrogate Markers of Injury Severity: A Clinical Validation Chamber Model, 273</td>
</tr>
<tr>
<td>Estabrooks, Carole A; Midodzi, William K; Cummings, Greta G; Ricker, Kathryn L; Giovannetti, Phyllis</td>
<td>The Impact of Hospital Nursing Characteristic on 30-Day Mortality, 74</td>
</tr>
<tr>
<td>Estabrooks, Carole A</td>
<td>jt auth See Cummings, Greta</td>
</tr>
<tr>
<td>Ferre, Lilian</td>
<td>jt auth See Peragallo, Nilda</td>
</tr>
<tr>
<td>Finkelnburg, Mary J; Wynman, Jean F; Crogan, Catherine F; Nyman, John A</td>
<td>Use of Time Studies for Determining Intervention Costs, 280</td>
</tr>
<tr>
<td>Finnegan, Lorna; Marion, Lucy; Cox, Cheryl</td>
<td>Profiles of Self-Related Health in Midlife Adults With Community Illnesses, 167</td>
</tr>
<tr>
<td>Fitzwater, Evelyn</td>
<td>jt auth See Gates, Donna Floy</td>
</tr>
<tr>
<td>Garrett-Mayor, Elizabeth</td>
<td>jt auth See Ulrich, Connie M</td>
</tr>
<tr>
<td>Gates, Donna; Fitzwater, Evelyn; Sucop, Paul A</td>
<td>Reducing Assault Nursing Home Caregivers, 119</td>
</tr>
<tr>
<td>Gennaro, Susan</td>
<td>jt auth See Amari, Angela F; Gennaro, Susan</td>
</tr>
<tr>
<td>Giddings, Lynne S</td>
<td>Health Disparities, Social Injusties, and the Culture of Nursing, 304</td>
</tr>
<tr>
<td>Giger, Joyce N</td>
<td>jt auth See Appel, Susan J; Giovannetti, Phyllis</td>
</tr>
<tr>
<td>Given, Barbara</td>
<td>jt auth See Doorenbos, Ardhith Z; Given, Barbara</td>
</tr>
<tr>
<td>Gomma, Lynda</td>
<td>jt auth See Ulrich, Connie M</td>
</tr>
<tr>
<td>Ham, Kuem Sun</td>
<td>jt auth See Hoe, Seongkum; Hammad, Tarek A</td>
</tr>
<tr>
<td>Han, Kuem Sun; Lee, Sook Ju</td>
<td>Park, Eun Sook; Park, Young-Joo; Cheol, Kang Hyun; Structural Model for Quality of Life Patients with Chronic Cardiovascular Disease in Korea, 85</td>
</tr>
<tr>
<td>Hanna, Tiffany</td>
<td>jt auth See Appel, Susan J; Harinrattisai, Teeranut; Johnson, Rebecca A</td>
</tr>
<tr>
<td>Havens, Donna S; Vasey, Joseph</td>
<td>The Staff Nurse Decision Involvement Scale: Report of Psychometric Assessments, 376</td>
</tr>
<tr>
<td>Hawkes, William</td>
<td>jt auth See Resnick, Barbara; Hayduk, Leslie</td>
</tr>
<tr>
<td>Hubbard, Ryan</td>
<td>jt auth See Ulrich, Connie M</td>
</tr>
<tr>
<td>Inguito, Pietro</td>
<td>jt auth See Resnick, Barbara</td>
</tr>
<tr>
<td>James, Dotti C; Jeffers, Brenda Recchia; Research Environments That Promote Integrative Care, 63</td>
<td></td>
</tr>
<tr>
<td>Johangten, Mary</td>
<td>jt auth See Lee, Sun-Mi; Johnson, Rebecca A</td>
</tr>
<tr>
<td>Kim, Young Ju</td>
<td>jt auth See Peragallo, Nilda; Kim, Young Ju; Soeken, Karen L; A Meta-Analysis of the Effect of Hospital-Based Case Management on Hospital Length-of-Stay and Readmission, 255</td>
</tr>
<tr>
<td>Kolanoowski, Ann M; Litaker, Mark; Bucetner, Linda; Efficacy of Theory-Based Activities for Behavioral Symptoms of Dementia, 219</td>
<td></td>
</tr>
<tr>
<td>Kosiol, Deloris</td>
<td>jt auth See Ulrich, Connie M</td>
</tr>
<tr>
<td>Lake, Eileen T</td>
<td>jt auth See Park, Sunhee Larson, Elaine L</td>
</tr>
<tr>
<td>Lerner, Mary H</td>
<td>jt auth See Teel, Cynthia S; Lin, Susan X;jt auth See Cimiotti, Jeanne P; Lin, Wen-Chieh</td>
</tr>
<tr>
<td>Magaziner, Jay</td>
<td>jt auth See Resnick, Barbara Manolovich, Milisa; Predictors of Professional Nursing Practice Behaviors in Hospital Settings, 41</td>
</tr>
<tr>
<td>Marek, Karen Dorman; Popejoy, Lori; Petrokis, Greg; Mehr, David; Rantz, Marilyn; Lin, Wen-Chieh; Clinical Outcomes of Aging in Place, 202</td>
<td></td>
</tr>
<tr>
<td>Marion, Lucy</td>
<td>jt auth See Finnegan, Lorna McDevitt, Judith</td>
</tr>
<tr>
<td>Mehr, David</td>
<td>jt auth See Marek, Karen D Midedzi, William K</td>
</tr>
<tr>
<td>Nyman, John A</td>
<td>jt auth See Findorff, Mary J O’Brien, Marian C</td>
</tr>
</tbody>
</table>
Singh, Jagdip  jt auth  See Punthmatharith, Busakorn
Smith, Andrew L  jt auth  See Dunbar, Sandra B
So, Winnie KW; Tai, Josepha WM; Fatigue and Fatigue-Relieving Strategies Used by Hong Kong Chinese Patients After Hemopoietic Stem Cell Transplantation, 48
Sosken, Karen L  jt auth  See Kim, Young-Ju
Steele, Susan K; Porche, Demetrious J; Testing the Theory of Planned Behavior to Predict Mammography Intention, 332
Stewart, Janet L; Lyn, Mary R; Mishel, Merle H; Evaluating Content Validity for Children's Self-Report Instruments Using Children as Content Experts, 414
Stone, Patricia W  jt auth  See Cimiotti, Jeannie P
Tabares, Amber  jt auth  See Carrère, Sybil
Tai, Josepha WM  jt auth  See So, Winnie KW
Teel, Cynthia S; Leenerts, Mary H; Developing and Testing Self-Care Intervention for Older Adults in Caregiving Roles, 193
Ulrich, Connie M; Danis, Marion; Kozioł, Deloris; Garrett-Mayer, Elizabeth; Hubbard, Ryan; Grady, Christine; Does It Pay so Pay?: A Randomized Trial Financial Incentives and Lottery Incentives in Surveys of Nonphysician Healthcare Professionals, 178
Valcante, Gregory  jt auth  See Elder, Jennifer H
Vasey, Joseph  jt auth  See Havens, Donna S
Vassallo, Annemarie  jt auth  See Wilbur, JoEllen
Verbitsky, Natalya  jt auth  See Doorenbos, Ardith Z
Vessey, Judith; Gennaro, Susan; The Gardener: Florence Downs August 20, 1925-September 8, 2005, 374
Wang, Chen-Yen; Chan, Siu Ming Alain; Culturally Tailored Diabetes Education Program for Chinese Americans: A Pilot Study, 347
Weaver, Michael T  jt auth  See Appel, Susan J
Werner, Michele  jt auth  See Resnick, Barbara
White, Deborah  jt auth  See Elder, Jennifer H
Whitemore, Robin; Combining Evidence in Nursing Research: Methods and Implications, 56
Wilbur, JoEllen; Vassallo, Annemarie; Chandler, Peggy; McDevitt, Judith; Miller, Arlene
Michaels; Midlife Women's Adherence to Home-Based Walking During Maintenance, 33
Woodin, Erica  jt auth  See Carrère, Sybil
Wooldridge, Powhatan J  jt auth  See Wu, Yow-Wu B
Wu, Yow-Wu B; Wooldridge, Powhatan J; The Impact of Centering First-Level Predictors on Individual and Contextual Effect in Multilevel Data Analysis, 212
Wyman, Jean F  jt auth  See Findorff, Mary J
Yahiro, Janet Yu  jt auth  See Resnick, Barbara
Yarandi, Hossein  jt auth  See Elder, Jennifer H
Yeh, Chao-Hsing; Chang, Cha-Wen; Chang, Pi-Chen; Evaluating Quality of Life in Children With Cancer Using Children's Self-Reports and Parent-Proxy Report, 354
Yoshimoto, Dan  jt auth  See Carrère, Sybil
Zimmerman, Sheryl  jt auth  See Resnick, Barbara