## Issue Table of Contents

January/February 2005, Volume 54, Issue 1

<table>
<thead>
<tr>
<th>Page</th>
<th>Article/Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>EDITORIAL</strong></td>
<td><strong>Research Thrust Changing? At Risk.</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Blegen, Mary A. PhD, RN, FAAN</td>
</tr>
<tr>
<td>2</td>
<td><strong>FEATURES</strong></td>
<td><strong>Mitigating the Impact of Hospital Restructuring on Nurses: The Responsibility of Emotionally Intelligent Leadership.</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cummings, Greta; Hayduk, Leslie; Estabrooks, Carole</td>
</tr>
<tr>
<td>13</td>
<td><strong>A Framework for Studying the Nutritional Health of Community-Dwelling Elders.</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chen, Cheryl Chia-Hui</td>
</tr>
<tr>
<td>22</td>
<td><strong>Evaluating In-Home Training for Fathers of Children With Autism Using Single-Subject Experimentation and Group Analysis Methods.</strong></td>
<td>Elder, Jennifer H.; Valcante, Gregory; Yarandi, Hossein; White, Deborah; Elder, Timothy H.</td>
</tr>
<tr>
<td>33</td>
<td><strong>Midlife Women's Adherence to Home-Based Walking During Maintenance.</strong></td>
<td>Wilbur, JoEllen; Vassalo, Annemarie; Chandler, Peggy; McDevitt, Judith; Miller, Arlene Michaels</td>
</tr>
<tr>
<td>41</td>
<td><strong>Predictors of Professional Nursing Practice Behaviors in Hospital Settings.</strong></td>
<td>Manojlovich, Milisa</td>
</tr>
<tr>
<td>48</td>
<td><strong>Fatigue and Fatigue-Relieving Strategies Used by Hong Kong Chinese Patients After Hemopoietic Stem Cell Transplantation.</strong></td>
<td>So, Winnie K. W.; Tai, Josepha W. M.</td>
</tr>
<tr>
<td>56</td>
<td><strong>METHODS</strong></td>
<td><strong>Combining Evidence in Nursing Research: Methods and Implications.</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Whittemore. Robin</td>
</tr>
<tr>
<td>63</td>
<td><strong>BRIEF REPORT</strong></td>
<td><strong>Research Environments That Promote Integrity.</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jeffers, Brenda Recchia</td>
</tr>
</tbody>
</table>
A strong statement asserting that research-based knowledge in nursing is of utmost importance should not surprise readers of this journal. During the half century since the first issue of *Nursing Research* was published, the growth in research supporting nursing practice has been phenomenal. The dynamic trajectory of nursing research through the last 20 years of the 20th century seemed to be a self-sustaining surge as more and more nurses were trained in research methods at the doctoral level and funding for nursing research was formalized at NIH with the creation of the NCNR, now the NINR. In these first years of the 21st century, however, there are signs that this forward thrust is slacking.

There is still an overwhelming need to produce knowledge that is valid, replicable, and generalizable, and on which we can base our practice interventions. As evidence-based practice becomes the accepted way to design nursing care, service settings have created mechanisms that support searching for the research pertaining to a clinical problem, synthesizing this research, and creating protocols to implement in practice. But for many nursing practice problems, there is no research on which to build protocols. For some nursing practice problems there are a small number of studies, some done using methods that were not meant to produce generalizable knowledge, and some intended to produce generalizable knowledge but contained sufficient threats to validity that made applying this knowledge questionable. Before our profession can move forward to evidence-based practice we must have better evidence.

And yet, within our profession there are movements away from rigorous research-based doctoral training. These movements are for the best intentions and respond to identified needs: more quickly produce nursing faculty in order to increase nursing school enrollments or to replace faculty members who are reaching retirement age; and to produce nurse leaders effective in the service settings to establish evidence-based practice and to influence public and corporate policy. However, well-intentioned movements that siphon off and soften the rigor of the research training that has been the foundation of doctoral education may, in the long run, be detrimental.

Without rigorous knowledge production, experts engaged in developing evidence-based practice will not have evidence to use, faculty teaching new nurses will not have content, and leaders promoting policies that support high-quality nursing care will not have arguments based in data. We must ensure that the limited resources available to support the future of our profession are used in a way that will continue to build the knowledge base.

Many barriers to nurses’ conducting research have been overcome by strong nurse researchers of the last several decades. We initially have earned credibility and respect from researchers in many disciplines and now must strive to maintain and enlarge that. This is not the time to back away from this challenge. We did not during the last 50 years and we must not during the next.

—Mary A. Blegen, PhD, RN, FAAN
Associate Editor
E-mail: Mary.Blegen@UCHSC.edu
Mitigating the Impact of Hospital Restructuring on Nurses

The Responsibility of Emotionally Intelligent Leadership

Greta Cummings ▪ Leslie Hayduk ▪ Carole Estabrooks

Background: A decade of North American hospital restructuring in the 1990s resulted in the layoff of thousands of nurses, leading to documented negative consequences for both nurses and patients. Nurses who remained employed experienced significant negative physical and emotional health, decreased job satisfaction, and decreased opportunity to provide quality care.

Objective: To develop a theoretical model of the impact of hospital restructuring on nurses and determine the extent to which emotionally intelligent nursing leadership mitigated any of these impacts.

Methods: The sample was drawn from all registered nurses in acute care hospitals in Alberta, Canada, accessed through their professional licensing body (N = 6,526 nurses; 53% response rate). Thirteen leadership competencies (founded on emotional intelligence) were used to create 7 data sets reflecting different leadership styles: 4 resonant, 2 dissonant, and 1 mixed. The theoretical model was then estimated 7 times using structural equation modeling and the seven data sets.

Results: Nurses working for resonant leaders reported significantly less emotional exhaustion and psychosomatic symptoms, better emotional health, greater workgroup collaboration and teamwork with physicians, more satisfaction with supervision and their jobs, and fewer unmet patient care needs than did nurses working for dissonant leaders.

Discussion: Resonant leadership styles mitigated the impact of hospital restructuring on nurses, while dissonant leadership intensified this impact. These findings have implications for future hospital restructuring, accountabilities of hospital leaders, the achievement of positive patient outcomes, the development of practice environments, the emotional health and well-being of nurses, and ultimately patient care outcomes.

Key Words: emotional intelligence • hospital restructuring • leadership • structural equation modeling

A decade of North American hospital restructuring in the 1990s resulted in the layoff of thousands of nurses, leading to documented negative physical and emotional health effects. Nurses who remained employed experienced significant reductions in job satisfaction and quality of care (Cummings & Estabrooks, 2003; Sochalski, 2001). Research results suggested that the negative impact on nurses carried over to patients in the form of reduced quality of care and increased patient mortality (Aiken, Clarke, Sloane, Sochalski, & Silber, 2002; Blegen, Goode, & Reed, 1998; Kovner & Gergen, 1998; Needleman, Buerhaus, Mattke, Stewart, & Zelevinsky, 2002). What remained unclear was whether all nurses experienced the effects of hospital restructuring to the same degree, or whether nurses working in environments that reflected emotional competence by the nursing leadership experienced reduced effects.

Daniel Goleman and his colleagues have written extensively on emotional intelligence (EI), recently asserting that while leadership attributes include analytic intelligence, task completion, and organizational skills, the primary role must extend to effectively responding to their own and others’ emotions (Goleman, Boyatzis, & McKee, 2002). They claimed that the most effective leaders were those with high EI, who portrayed resonant leadership. This leadership reflects the art of hearing their workers’ negative feelings yet responding empathically. In times of change and even chaos, an effective leader needs to be empathic and supportive, and demonstrate a wide range of EI competencies (Goleman et al., 2002). Empathy, or the ability to comprehend another person’s feelings and to
reexperience them oneself, has been reported as a central component of EI (Salovey & Mayer; 1990) and the key to successful resonant leadership (Goleman et al., 2002). Empathic leaders are attuned to a wide range of emotional signals, allowing them to sense the felt, but unspoken, emotions in another person or group (Goleman et al., 2002; Wolff, Pescosolido, & Druskat, 2002).

The current study sought to determine the extent to which EI (resonant) leadership mitigated the detrimental consequences of hospital restructuring to nurses.

**Relevant Literature and Research**

**Leadership and Emotional Intelligence**

Emotionally intelligent leaders inspire by engaging emotions, passions, and motivations that reveal the possibility of achieving goals that might not otherwise be seen. They work through emotion to mobilize teams, coach performance, inspire motivation, or create a vision for the future. Goleman and colleagues’ (2002) view of EI has been based on four domains, self-awareness, self-management, social awareness, and relationship management, each consisting of several competencies. The self-awareness and self-management domains reflect personal competence in understanding and managing personal emotions. Social competence (the ability to develop and manage relationships with others) is composed of social awareness and relationship management domains. Goleman and colleagues reframed these EI competencies to reflect leadership competencies, using them to describe and distinguish six leadership styles. Four leadership styles (visionary, coaching, affiliative, and democratic) were termed resonant because they demonstrated high levels of EI, and two styles (pace setting and commanding) were dissonant because they failed to demonstrate EI. Resonant leaders’ messages are tuned to their own and others’ feelings as they build harmony and positive working climates. Dissonance, an unpleasant, harsh experience in both musical and human terms, references a lack of harmony and being emotionally “out of touch” with employees. Dissonant leadership undermines the emotional foundations that support and promote staff success (Goleman et al., 2002).

In this study, Goleman and colleagues’ EI competencies were matched to theoretical descriptions of their six leadership styles. For example, Goleman and colleagues (2002) described visionary leadership as being able to move people toward shared dreams, having empathy for and developing relationships with others, sharing knowledge to empower others to innovate, having integrity (transparency), and continually reminding people of the greater purpose of their work. These characteristics were required to be present before a nurse in this study was classified as working in a visionary leadership environment. Likewise, as pace-setting and commanding leadership is not known for empathy or developing others, these latter characteristics were required to be absent in environments defined by dissonant leadership. Goleman and colleagues’ descriptions of coaching, affiliative, and democratic styles also informed the determination of environments reflecting their requisite characteristics. Coaching leadership focuses primarily on developing others and achieving high levels of individual performance, affiliative leadership builds strong relationships, and democratic leadership builds consensus and promotes innovation, teamwork, and collaboration.

Whether the most appropriate means of assessing leadership competence is through direct ability testing, self-assessment, or by the perceptions of those actually working for the leader has been discussed broadly. Most researchers have concluded that having actual workers rate their leaders provides the best construct validity (Bass & Avolio, 1991; Dasborough & Ashkanasy, 2002; Dunham, 2000; Kellett, Humphrey, & Sleeth, 2002; Xin & Pelled, 2002). This tradition was followed when nurses’ responses to specific survey questions were analyzed in this study.

**Moderating or Mitigating Action of Leadership**

While considerable literature has supported the notion that EI contributes to effective leadership (Freshman & Rubino, 2002; McColl-Kennedy & Anderson, 2002; Robbins, Bradley, Spicer, & Mecklenburg, 2001; Snow, 2001), the literature on leadership as a moderator of effects has reported mixed results (de Vries, Roe, & Taillieu, 2002; Gavin & Hofmann, 2002; Pirola-Merlo, Haertel, Mann, & Hirst, 2002; Villa, Howell, Dorfman, & Daniel, 2003). Research on the role and responsibility of leadership in mitigating (lessening in force or intensity) the consequences of massive organizational change to employees was not found. Recent doctoral studies investigating EI and leadership among nurses have not considered leadership as partially mitigating the effects of hospital restructuring on nurses (Graves, 2000; Molter, 2002; Tjong, 2002; Vitello-Cicciu, 2002).

A decade of restructuring and downsizing placed prolonged pressures on the nurses of Alberta and provided an opportunity to examine whether various styles of leadership differentially protected nurses from the effects of budget cutting.

**Methods**

**Survey and Sample**

The Alberta Nurse Survey of Hospital Characteristics (Giovannetti, Estabrooks, & Hesketh, 2002), one provincial component of the Canadian portion of the International Survey of Hospital Staffing and Organization of Patient Outcomes (Aiken et al., 2001), was used for this analysis. The survey, completed in 1998, reported on various organizational attributes and the state of Alberta nurses’ physical and emotional well-being (Giovannetti et al., 2002). The survey had 139 questions in seven categories (employment characteristics, nursing work index, burnout inventory, staffing, details of the last worked shift, quality of care, demographic characteristics), and questions that addressed restructuring, violence in the workplace, and the use of information resources (Giovannetti et al., 2002). Alberta was the second Canadian province to undergo regionalization of its health authority structures in the mid-1990s, collapsing 283 hospital and health care boards into 17 regional health authorities. This, along with

---

1Additional information provided by the authors expanding this article can be found at the editor’s Web site (http://nursing-research-editor).
other initiatives in health care reform, led to the layoff of thousands of nurses in Alberta throughout the 1990s (Maurier and Northcott, 2000). The Alberta portion of the Canadian Nurse Survey was used because it was the only province in which specific questions were asked about local hospital restructuring, including how many times each nurse was laid off or was required to change nursing units as a result of hospital restructuring.

The study sample was drawn from all registered nurses working in acute care hospitals in Alberta, Canada. Nurses were invited to participate through their professional licensure association. The final sample included 6,526 nurses (53% response rate); demographic comparisons showed no significant difference from the acute care nursing population (Giovannetti et al., 2002).

Data Sets
Seven data sets were created, each consisting of data from nurses who practiced in environments reflecting different leadership styles. There were 13 questions chosen from the nurse survey to reflect EI leadership competencies and sort each case into one or more data sets. Thirteen leadership competencies were used because few survey questions captured how leaders managed their own emotions (personal competence). Goleman and colleagues (2002) argued that only six to eight EI competencies were common even for successful leaders. Each of our resonant and dissonant leadership styles was defined by six to eight competencies. The required presence or absence of each competency was determined to fit each leadership style. Nurses had indicated the degree to which each particular statement described their current work environment using a 4-point Likert-type scale (from strongly agree to strongly disagree). By reporting on the presence or absence of a variety of work environment features, nurses had provided information that identified the styles of their nursing leaders. Therefore, presence of a leadership competency in the nurse’s work environment was identified by the nurse’s response of “agree” or “strongly agree” that the specific statement described their work environment. A nurse’s response of “disagree” or “strongly disagree” was taken as absence of evidence of that leadership competency in their work environment. A nurse’s survey data were included in the data set describing a specific style of leader (eg, visionary) if that nurse reported both the presence of all the leadership characteristics required by that style and the absence of all the characteristics contraindicating that style. The cases that fit the four resonant styles were selected out of the main data set and transferred into the four resonant data sets. Then the cases that fit the two dissonant styles were selected out and all remaining cases were placed into the Mixed Leadership Styles data set. Leadership Empathy, as indicated by “administration that listens and responds to employee concerns,” was the competency that differentiated the dissonant and resonant groups.

Information on the supervisory environments was not complete enough to unambiguously classify the leadership environment of some nurses as reflecting only one leadership style. However, consistent with Goleman and colleagues’ concepts, leaders portray different styles depending on the situation at hand—visionary when inspiration is called for and democratic when consensus team building is needed. Therefore, if a nurse’s response pattern was compatible with the characteristics of two different leadership styles, that nurse’s data was included for analysis in both data sets. Such multiple classifications appeared only within the resonant leadership styles or within the dissonant styles, not between them. Fortunately, the major differences in outcomes appear between the three distinct groups of resonant, dissonant, and mixed leadership styles; therefore cross-classifications between the groups did not confound the important results. Differences in the means and standard deviations for each variable across all seven styles confirmed that each data set reflected a different population of nurses. The sample sizes for each of the six leadership styles ranged from 699 to 1,063. Nurses who worked in environments that reflected leadership styles other than the four resonant or two dissonant styles were placed together in a seventh group of “mixed” leadership styles. Initially, over half (N = 3,868) of the entire Alberta survey data set were classified into this group. In order to ensure that the statistics sensitive to sample size were at least relatively comparable, 1,065 cases were randomly selected from the mixed leadership data set for analysis.

Model Development
A theoretical model was developed that portrayed causal relationships between hospital restructuring (background causal variables) and effects on nurses (outcome variables), and that used the results of a systematic review of the research literature to determine these relationships (Cummings & Estabrooks, 2003). The causal relationships among the outcome variables were derived from the literature and the primary author’s leadership experience during hospital restructuring in a large tertiary care hospital. The key elements of the model are described below, and the theoretical model underlying the research is posted at the Editor’s Web page (http://nursing-research-editor).

Causal Variables (Hospital Restructuring) The causal variables included the number of restructuring events occurring in the hospital, being laid off in the past 5 years, changing units in the past 5 years, along with several demographic variables (years worked in a hospital, part-time/full-time status, and age). Gender was included as a
control response variable, that is, as having no effect. Summing positive responses to seven questions that asked whether specific hospital restructuring events, such as loss of the senior nursing position without replacement or an increase in the number of patients assigned to each nurse, had occurred in their hospital was used to derive the number of restructuring events experienced by each nurse. The number of times that each nurse reported being laid off or changing units, the number of years worked in that hospital, and the nurse’s age were entered as reported. Work status was coded 1 for part-time and 2 for full-time.

**Nursing Outcome Variables** The nursing outcome variables included nurse reports of freedom to make important patient care decisions, emotional health, satisfaction with time to spend with patients, teamwork between physicians and nurses, nursing workgroup collaboration, satisfaction with supervision, satisfaction with financial rewards, job mobility options, job security, and job satisfaction. The degree to which each feature was perceived to be present in the respondent’s workplace had been answered on a 4-point Likert-type scale. Intent to quit was measured on a 3-point scale. Nurses had recorded their degree of emotional exhaustion, psychosomatic symptoms, and professional efficacy on 6-point scales from never (0) to every day (6). An important variable, unmet patient care needs (a proxy measure for quality of care used by Sochalski [2001]), was derived by summing the number of patient care tasks (maximum of 8) that were deemed necessary by the nurse but were left unattended by the end of the last worked shift.

Each concept in the theoretical model was indexed to a single indicator from the nurse survey. On the basis of the authors’ judgment of how accurately the specific indicator reflected the corresponding underlying latent concept, an adjustment was made for the quality of each indicator by assigning 2–30% of its variance as error. The percentages of measurement error were determined by carefully examining how closely each latent variable in our theoretical model was being measured by its indicator in the data sets. Thus, a compensation for problematic wordings, lack of clarity in some questions, and other measurement concerns was made. Pairwise covariance matrices were created because listwise deletion would have resulted in the loss of too many cases.

**Model Estimation and Testing Results** The same model was estimated (using Lisrel 8.20 maximum likelihood estimation; Jöreskog & Sörbom, 1996) for each of the seven leadership style data sets. The chi-square ($\chi^2$) for the seven models ranged from 205.21 ($p < .001$) to 340.26 ($p < .001$) and the adjusted goodness of fit index (AGFI) ranged from .928 to .945, indicating substantial inconsistencies between the models and the data sets (Hayduk, 1987). The seven models were examined carefully to locate model modifications that were theoretically tenable and that could be made uniform across the seven leadership models. Three criteria were used: the change had to be reasonable theoretically; the modification indices for the relevant coefficients had to be greater than 7 in three or more models, or greater than 10 in two or more models; and reciprocal effects that would have resulted in underidentified models were avoided. Modifications were applied to all models, not merely those with the substantial modification indices. This consistency reduced the likelihood of capitalizing on chance sampling fluctuations that might have existed across the seven data sets. If an effect truly was not required, it would merely lead to a null coefficient estimate and hence would not harm the model, but would cost a degree of freedom. Seventeen additional coefficients were added to the model for estimation using these decision rules.

The large decrease in chi square after these modifications (Table 1) indicated a substantially improved, but not completely acceptable, model fit (Jöreskog & Sörbom, 1996; Hayduk, 1987). The risk of biased estimates due to model misspecification had to be balanced against the risk of bias resulting from inserting “effects” corresponding to chance sampling fluctuations in the covariances. The requirement of inserting changes consistent across the seven models had provided substantial protection against bias due to sampling-fluctuation–induced controls during the numerous changes that had already been made, but the risk of improper-control biases would have increased markedly had unique changes to each model been applied. Most of the models were only one or two modifications away from chi-square fit, but the likely small size of the estimates and the markedly increased risk of improper-control bias led the authors to believe that the estimates in the slightly ill-fitting models were the best estimates attainable.

**Analysis** The analysis occurred in two stages. Initially, the estimated coefficients for the effects of hospital restructuring on nursing outcomes for each leadership style (effects within each leadership group) were analyzed. Then the impact of leadership styles on the nursing outcome variables (i.e., differences among leadership groups) was analyzed.

**The Impact of Hospital Restructuring on Nursing Outcomes** The direction and significance of 50 effects of hospital restructuring on nursing outcomes are reported in Table 2. The results of model estimation included 16 effects that were significant, of which 6 were significant in all seven leadership styles. The two largest and most significant effects were found in seven leadership styles: (a) the direct relationship between the number of hospital restructuring events and the reported number of patient care needs left unattended, and (b) the direct relationship

---

1. See Footnote 1.
2. The covariance matrices for seven leadership styles may be obtained from the primary author at gretac@ualberta.ca.
3. Four coefficients in the original model that satisfied the first decision rules but that would have created loops were not acted upon.
TABLE 1. Fit of the Initial and Final Models

<table>
<thead>
<tr>
<th>Theoretical model</th>
<th>Visionary</th>
<th>Coaching</th>
<th>Affiliative</th>
<th>Democratic</th>
<th>Mixed</th>
<th>Pace Setting</th>
<th>Commanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significance</td>
<td>( p &lt; .001 )</td>
<td>( p &lt; .001 )</td>
<td>( p &lt; .001 )</td>
<td>( p &lt; .001 )</td>
<td>( p &lt; .001 )</td>
<td>( p &lt; .001 )</td>
<td>( p &lt; .001 )</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>113</td>
<td>113</td>
<td>113</td>
<td>113</td>
<td>113</td>
<td>113</td>
<td>113</td>
</tr>
<tr>
<td>AGFI</td>
<td>0.934</td>
<td>0.939</td>
<td>0.945</td>
<td>0.945</td>
<td>0.937</td>
<td>0.928</td>
<td>0.933</td>
</tr>
<tr>
<td>Following 17 modifications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chi square</td>
<td>129.717</td>
<td>138.804</td>
<td>122.625</td>
<td>155.885</td>
<td>205.642</td>
<td>147.318</td>
<td>175.487</td>
</tr>
<tr>
<td>Significance</td>
<td>( p = .012 )</td>
<td>( p = .002 )</td>
<td>( p = .034 )</td>
<td>( p &lt; .001 )</td>
<td>( p &lt; .001 )</td>
<td>( p &lt; .001 )</td>
<td>( p &lt; .001 )</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>96</td>
<td>96</td>
<td>96</td>
<td>96</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>AGFI</td>
<td>0.957</td>
<td>0.962</td>
<td>0.966</td>
<td>0.966</td>
<td>0.955</td>
<td>0.949</td>
<td>0.949</td>
</tr>
<tr>
<td>N</td>
<td>699</td>
<td>851</td>
<td>716</td>
<td>1,065</td>
<td>1,065</td>
<td>674</td>
<td>799</td>
</tr>
</tbody>
</table>

Note. AGFI = adjusted goodness-of-fit index.

TABLE 2. Significant and Nonsignificant Effects of Hospital Restructuring Variables on Nursing Outcomes

<table>
<thead>
<tr>
<th>Causal Variable</th>
<th>Nursing Outcome Variable</th>
<th>Hospital Restructuring Events</th>
<th>Times Nurse Changed Units</th>
<th>Part-Time/Full-Time Status</th>
<th>Years of Experience in Current Hospital</th>
<th>Times Nurse Laid Off</th>
<th>Age</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Unmet patient care needs</td>
<td>+</td>
<td>ns</td>
<td>ns</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Freedom to make important patient care decisions</td>
<td>ns</td>
<td>ns*</td>
<td>ns*</td>
<td>+/—</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Professional efficacy</td>
<td>ns*</td>
<td>ns*</td>
<td>ns*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Satisfaction with time to spend with patients</td>
<td>ns*</td>
<td>+/—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Emotional exhaustion</td>
<td>+</td>
<td>ns</td>
<td>+*</td>
<td>ns</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Psychosomatic symptoms</td>
<td>ns</td>
<td>ns*</td>
<td>ns*</td>
<td>ns*</td>
<td>-*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Emotional health</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Teamwork between physicians and nurses</td>
<td>ns*</td>
<td>ns</td>
<td>ns*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Nursing workgroup collaboration</td>
<td>—</td>
<td>ns*</td>
<td>ns*</td>
<td>ns</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Job security</td>
<td>ns</td>
<td>ns*</td>
<td>+*</td>
<td>—*</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Satisfaction with supervision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Satisfaction with financial rewards</td>
<td>ns*</td>
<td></td>
<td></td>
<td>ns</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Job mobility options</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Job satisfaction</td>
<td>ns*</td>
<td>ns*</td>
<td>+/—</td>
<td>ns*</td>
<td>ns*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Intent to quit</td>
<td>ns*</td>
<td>ns</td>
<td>ns*</td>
<td>—*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. + = A significant positive effect was estimated in at least two of three leadership-style groups (resonant, dissonant, and mixed). — = A significant negative effect was estimated in at least two of three leadership-style groups (resonant, dissonant, and mixed). ns = Estimated effect was not significant in all, or at least most, of the leadership styles. * = Effect was consistent across all seven leadership styles. Blank = Effect was not estimated.
between full-time status and increased emotional exhaustion in nurses. Greater numbers of hospital restructuring events also led to nursing reports of greater emotional exhaustion, deterioration in emotional health, and disruption to workgroup collaboration. The only differential effects between full-time and part-time nurses were that full-time status not only led to more emotional exhaustion but also more satisfaction with their job and time to spend with patients, than did part-time status. The number of times that nurses changed units had no significant effect, and the number of times that nurses were laid off resulted only in the perception of decreased job security.

The direction and significance of 63 causal relationships among the outcome variables are reported in Table 3. Thirty-nine effects were significant, and, of these, nine were significant in all leadership styles. Emotional exhaustion led to the greatest number of significant outcomes, including more psychosomatic symptoms and unmet patient care needs, and deterioration in emotional health, satisfaction with financial rewards, and job satisfaction. Job security and satisfaction with supervision improved nurses’ emotional health. Professional efficacy and freedom to make important patient care decisions led to fewer patient care needs being left unattended. The more the patient care needs that were not met, the lower the nurses’ satisfaction with the time to spend with patients, which further reduced nursing workgroup collaboration. Freedom to make important patient care decisions enhanced teamwork between physicians and nurses, leading to increased job satisfaction.

### Impact of Leadership Styles

Determining the impact of specific leadership styles involved examining the degree to which nurses experienced the consequences of hospital restructuring depending on the leadership style characterizing their work environment. This analysis was enabled by graphing each leadership style’s effect coefficient with the means of the two variables contributing to that particular effect (Figure 1). Figure 1a is discussed in some detail and the remainder is summarized. The slope of each of the seven lines depicted in Figure 1a is the estimated effect coefficient of hospital restructuring events on unmet patient care needs for each leadership style. These show that hospital restructuring led to reported increases in unmet patient care needs for all nurses surveyed. However, the placement of each line is determined by the means of both variables (hospital restructuring events and unmet patient care needs) for each specific leadership style and illustrates the degree to which the outcome variable (unmet patient care needs) differed by leadership style. Therefore, the differences between the seven lines are the impact of leadership styles as reported by nurses on the relationship between hospital restructuring and unmet patient care needs. As the number of hospital restructuring events increased, so
FIGURE 1. Selected examples of the impact of leadership styles to mitigate or intensify the consequences of hospital restructuring on nurses. ♦, Visionary; ■, Coaching; ▲, Affiliative; †, Democratic; ★, Unknown; •, Pacesetting; ●, Commanding.
Discussion

In this study, leadership was reflected through EI. Although these findings are diverse, three areas are highlighted further: leader–staff relationships, implications for organizational policy, and study limitations that should guide additional research.

Leader–Staff Relationships

The results of this theoretical model estimation showed that all nurses felt the effects of hospital restructuring; however, nurses who worked in resonant leadership environments reported fewer negative effects. This would lead to greater satisfaction and more “emotional resilience” with which to provide quality care. This was evidenced by fewer necessary patient care needs being left unattended. The current findings suggest that resonant leaders used their emotional skills to understand what individual employees or teams were feeling during difficult times, thereby building trust through listening, empathy, and responding to staff concerns. The results suggest that after layoffs, resonant leaders be expected to work with remaining staff to understand their issues, their increased workload and emotional turmoil resulting from the layoff of colleagues and changes in practice patterns. After hospital restructuring, the results suggest these resonant leaders also would continue to invest in staff development and to consider nurses’ freedom to make important patient care decisions a high priority.

The results suggest that dissonant leadership would not be tuned to staff members’ emotional needs or focus on developing or maintaining relationships with them. Leaders exhibiting mixed leadership styles may be perceived as somewhere between the resonant and dissonant leadership styles. Also, these leaders may demonstrate a resonant style in one situation and a dissonant style in another. The large standard deviations in the means of the mixed leadership style group suggest that the latter was the case, and at minimum that the EI behaviors of this undifferentiated leadership group were diverse.

Study Limitations

We examined the translation validity (Trochim, 2003) of our process to use Goleman and colleagues’ description of six leadership styles into our research. Many of the 13 questions used referred to specific behaviors of the nurse’s supervisor or manager, adding to construct validity. Other questions were chosen on the basis of the primary author’s judgment in collaboration with an external expert in clinical outcome and health system research and the third author. To ensure that each question appropriately reflected a specific EI competency, only questions that described a specific characteristic that is a frontline leader’s responsibility and that could be perceived by a staff nurse as an EI competence were selected. Concurrent validity (Trochim, 2003) of the three theoretical groupings of leadership styles (resonant, mixed, and dissonant) was supported by using the required presence or absence of 13 EI leadership competencies to sort cases into one of the three mutually exclusive groups. The significant difference in study results for each of the three groupings of leadership styles suggests that we had achieved discrimination between these three groupings.

Future research into the mitigation of the consequences of adverse events should be done prospectively, and by using instruments designed to measure the EI of leaders as perceived by followers. It is recommended also that future research examine the influence of both the nursing manager’s
FIGURE 2. Selected examples of the impact of leadership styles to mitigate relationships among the nursing outcome variables. ♦, Visionary; ■, Coaching; ▲, Affiliative; ×, Democratic; ∧, Unknown; ●, Pacesetting; †, Commanding.
and the senior nursing leader’s styles on nurses and on the nursing work environment.

**Implications for Organizational Policy**

The findings suggest that resonant leadership would result in better quality of care by frontline providers. Hiring resonant leaders or providing training for existing leaders hence becomes a priority consideration for chief executives and nursing administrators, albeit recognized that screening for and assessing EI competencies in the workplace is still problematic (Matthews, Zeidner, & Roberts, 2002). The findings also suggest that nurses who reported characteristics of resonant leadership also reported enhanced teamwork between physicians and nurses, nursing workgroup collaboration, and the freedom to make important patient care decisions—all of which are important aspects of nursing practice environments. Health Canada (2000) has advocated that healthcare organizations reduce their occupational health and financial risk by establishing an organizational climate of fairness, purpose, and trust, in which staff wellness is a priority, leading to greater staff satisfaction. The findings suggest that resonant nursing leadership is a key—but missing—ingredient in this model for achieving these goals in hospitals.

Researchers and developers of EI training programs have identified that it is possible to learn how to increase EI. Earlier studies have shown that by wanting to learn and choosing to make a sustained, intentional behavioral change, people can change their performance on a complex set of competencies that distinguish outstanding managers (Boyatzis, 2001). The Consortium for Research on EI has summarized empirical findings on the best mode of learning EI competencies and has guidelines for developing training programs (Goleman et al., 2002). The incorporation of EI training into the basic nursing curriculum has also been identified as essential for nursing education (Evans & Allen, 2002).

It has not been implied that by employing resonant leaders, hospitals can mitigate—and thereby justify—the adverse effects of restructuring. Resonant leadership in this study did not eliminate the negative effects of hospital restructuring on nurses; however, they did lessen some of the negative effects that resulted and did so to a greater extent than dissonant or mixed leadership.

A theoretical model of causal relationships between hospital restructuring events and negative consequences to nurses’ work and health was developed and tested. These findings indicated that numerous detrimental effects to nurses' health and ability to provide quality care to patients resulted from widespread changes to hospitals. Nurses who experienced negative consequences of restructuring and worked in resonant leadership environments experienced these effects to a much lesser degree than those who worked under the influence of dissonant leadership. Resonant leadership mitigated most of the effects of hospital restructuring on their nurses, while dissonant leadership intensified these same effects. These findings suggest that by investing energy into relationships with nurses, resonant nursing leaders positively affect the health and well-being of their nurses, and, ultimately, the outcomes for patients.

**References**


Accepted for publication August 25, 2004.

The authors acknowledge the Canadian Institutes for Health Research (CIHR), the Alberta Heritage Foundation for Medical Research (AHFMR), and the University of Alberta for doctoral fellowships to Dr. Cummings that supported this work and the CIHR career award support for Dr. Estabrooks.

Corresponding author: Greta Cummings, PhD, RN, Faculty of Nursing, University of Alberta, 3-120 Clinical Sciences Bldg, Edmonton, Alberta, Canada T6G 2G3 (e-mail: greta@ualberta.ca).


A Framework for Studying the Nutritional Health of Community-Dwelling Elders

Cheryl Chia-Hui Chen

Background: Ensuring nutritional health for elders has been shown to reduce healthcare costs and enhance quality of life. Studies, however, have shown that malnutrition is present in 2% to 51% of community-dwelling elders, depending on the definition used and the population studied. An empirically tested framework for studying nutritional health in community-dwelling elders is not yet available.

Objective: To test the goodness-of-fit exhibited by the framework of nutritional health among community-dwelling elders based on the Roy Adaptation Model using structural equation modeling (SEM).

Methods: A population-based study investigated 243 elders dwelling in public housing. Demographics, polypharmacy, chronic illness, oral health, depressive symptoms, functional status, and satisfaction with social support were assessed to test their relation with nutritional health according to the propositions of the Roy Adaptation Model and scientific evidence.

Results: The SEM analysis indicated that functional status, oral health, depressive symptoms, and satisfaction with social support affect nutritional health directly. Oral health, depressive symptoms, functional status, and satisfaction with support mediated the effects of age, ethnicity, education, and number of medications and chronic illnesses on nutritional health. The model accounted for 35% of the variance in nutritional health and demonstrated a good fit with the data and with the values for Bentler's Comparative Fit Index (0.94) and \( \chi^2 \) (1.76).

Conclusions: The propositions of the Roy Adaptation Model were supported, and the findings showed that this framework of nutritional health among community-dwelling elders could serve as a theoretical and empirical base for future inquiry.

Key Words: aged • nursing theory • nutrition

Malnutrition is a frequent and serious problem among elders. In the United States, it is estimated that 40% of nursing home residents, 50% of hospitalized elders, and 45% of homecare elders are malnourished (Nutrition Screening Initiative, 1993). More than 85% of community-dwelling elders have one or more chronic illnesses that could be improved with proper nutrition, yet many of them suffer from various forms of malnutrition (Millen, 1999).

Little research has been directed toward determining the etiology of nutritional health among community-dwelling elders. Problems related to old age, functional status, oral health problems, chronic illness, polypharmacy, poverty, social isolation, and depressive symptoms have had a negative impact on nutritional health (Chen, Schilling, & Lyder, 2001). Yet some of these factors, such as social support, have not been empirically tested. Most studies also have been limited to examining bivariate relations, so it is unclear which factors are most important given their high correlation.

Understanding the etiology of nutritional health could assist nurses in developing targeted interventions for preventing and treating malnutrition. Therefore, an empirically tested framework of nutritional health among community-dwelling elders is needed to guide such efforts. The purpose of this study was to test the goodness-of-fit exhibited by the framework for studying nutritional health of community-dwelling elders based on the Roy Adaptation Model.

Review of the Literature

The success of nutrition care depends on careful identification of those elders at greatest risk for malnutrition so
Nutritional Health of Community-Dwelling Elders

Effectual interventions can be delivered. Such efforts must take into account the diversity and complexity of nutritional problems, the physiologic and psychosocial factors that may predispose an elder to malnutrition, and the adverse effects of preexisting and incident illnesses on nutritional health. On the basis of this theoretical proposition, the Roy Adaptation Model (RAM) was used as a conceptual framework for a holistic understanding of nutritional health among community-dwelling elders.

The RAM depicts humans as biopsychosocial beings that must adapt to the environmental stimuli. Problems arise when the individual's adaptive system is unable to cope with constantly changing stimuli from the environment in a manner that maintains the integrity of a system. Environmental stimuli are categorized as focal (stimuli immediately confronting individuals), contextual (contributing factors), and residual (unknown factors that may influence the situation). Adaptation takes place in four interrelated biopsychosocial response modes: physiologic, self-concept, role function, and interdependence. Biopsychosocial responses in any one mode may act as stimuli for input into the system (Roy & Andrews, 1999). Human health is a function of input stimuli and the individual adaptation level. Nutritional health, therefore, is conceptualized as an important aspect of the adaptation level. Environmental stimuli include demographics and the number of chronic illnesses. Biopsychosocial responses that can be observed from the four interrelated modes are hypothesized to include oral health, polypharmacy, depressive symptoms, satisfaction with support, and functional status (Figure 1) according to the propositions of RAM and scientific evidence.

Environmental Stimuli

Demographics
- Age has been suggested as one of the most important factors for nutritional health. Physiologic alterations resulting from the course of aging have been linked with an augmented probability of anorexia and further development of malnutrition (Morley, 1997). Evidence also exists that losses in sensory function and change in body composition increase the risk of malnutrition for elders. Additionally, low socioeconomic status has been widely accepted as a risk factor of malnutrition (Miller & Morley, 1995). The research findings, however, are inconclusive. Living alone and other sociodemographics did not adversely affect nutritional health in a cross-country sample of community-dwelling European elders (Pearson, Schlettwein-Gsell, Van Staveren, & de Groot, 1998).

Number of Chronic Illnesses
- Worldwide, malnutrition is most often associated with a lack of food. In developed countries, illnesses are the main contributor to malnutrition. For elders with heart disease, anorexia may be a part
of the cause for cardiac cachexia. Parkinson's disease and rheumatoid arthritis can induce malnutrition by causing hypermetabolism, anorexia, swallowing difficulty, and malabsorption. Chronic obstructive pulmonary disease causes weight loss in up to 71% of patients (Thomas, 1999). Malnutrition is the most common secondary diagnosis among cancer patients. Nutrition can be compromised by the systemic effects of the neoplasm and an anatomic location of the cancer that results in a mechanical barrier to eating. The physical and emotional stress from a cancer diagnosis and treatment can further suppress appetite and lead to poor food intake (Cunningham & Bell, 2000). Diabetes is associated with decreased zinc absorption and hyperzincuria. Zinc deficiency is associated with anorexia and poor immune function. Findings also have shown that elders with diabetes are at a higher risk for increasing susceptibility to periodontal disease, an increasing caries rate, a prolonging of fungal and bacteria infections, xerostomia, numbness, burning sensations, oral pain, and altered taste sensations (Lamey, Darwazeh, & Frier, 1992). All these may have negative effects on nutritional health.

**Biopsychosocial Responses of Adaptation Modes**

**Physiologic Mode: Oral Health and Polypharmacy** Oral health and polypharmacy represented the threats to physical integrity in this study. Elders who experience mouth pain, chewing or swallowing difficulties, poor dentition, ill-fitting dentures, or dry mouth that makes eating uncomfortable are at risk for malnutrition (Saunders, 1997). A Belgium study showed that the risk of malnutrition increases with a loss of natural teeth and the wearing of dentures (Cripe, Mets, Collys, Ponjaert-Kristoffersen, & Massart, 2000). Poor oral health may contribute also to embarrassment and avoidance of social interaction. An Italian study showed that dental status was a predictor of satisfactory social relationships for 1,137 community-dwelling elders (Appollonio, Carabellese, Frattola, & Trabucchi, 1997).

Community-dwelling American elders take a daily average of 2.7 to 4.2 prescriptive and over-the-counter medications (Hanlon, Schmader, Rudy, & Weinberger, 2001). Polypharmacy, defined as the concomitant use of medications, is a common problem among elders (Lyder, Fennie, Chen, & Fulmer, 2000). The risk of malnutrition induced by drugs is a major consequence of polypharmacy. Mechanisms of drug-nutrient interactions include reduced intake caused by side effects such as anorexia, nausea, and altered taste perception. Furthermore, medications may interfere with nutrient absorption, cause alteration in nutrient metabolism, and increase nutrient excretion (Roe, 1994). Measured as the number of medications taken, polypharmacy was shown to be one of the strongest predictors for nutritional health among retired, home-living elders (Cripe et al., 2000).

**In the United States, it is estimated that 40% of nursing home residents, 50% of hospitalized elders, and 45% of homecare elders are malnourished**

**Self-Concept Mode: Depressive Symptoms** The basic human need within the self-concept mode is psychic integrity (i.e., awareness of who one is that enables existence with a sense of unity) (Roy & Andrews, 1999). Depressive symptoms composed the self-concept mode in this study and have been linked to weight loss. In elders with depressive symptoms, approximately 90% lose weight, as compared with 60% of their younger counterparts. One American study identified depressive symptoms as the most common cause of malnutrition in a sample of elderly outpatients (Wilson, Vaswani, Liu, Morley, & Miller, 1998).

**Interdependence Mode: Social Support** Interdependence is a social adaptive mode, for which satisfying affectionate relationships with significant others and the provision and receipt of social support have been emphasized (Roy & Andrews, 1999). Studies have shown that an increase in social interaction at meal times improves dietary intake (de Castro, Brewer, Elmore, & Orazco, 1990). Social support also has been linked to mortality. Blazer (1982) found that social support was a significant factor in a 30-month mortality period. However, the significance of relationships depended on the definition of support. A measure of satisfaction with support was more related to mortality than support, defined either as a frequency of interaction or as availability of roles and attachment (e.g., marital status, number of children).

This finding is consistent with that of other studies in nutrition literature. One study showed that elders who belonged to social clubs did not have better nutritional status (Posner, Jette, Smigelski, Miller, & Mitchell, 1994). Also it was reported that the number of social contacts does not correlate with dietary intake (Walker and Beauchene, 1991).

**Role Function Mode: Functional Status** The role function mode is concerned with the individual’s preference of role activities on the basis of his or her position within society. Each role involves expected behaviors. Functional status, therefore, was assessed within the context of the role function mode. Functional status, defined as activity of daily living, has emerged as a risk factor of malnutrition in some studies (Guralnik & Simonsick, 1993). Although physical competence is the center of attention for most studies, social competence (e.g., the loss of social interest and role activity that accompanies retirement) and the death of family or friends, should not be discounted.

Inability of elders to maintain their social competence because of diminished physical ability is one aspect that has not been explored. Using the Enforced Social Dependency Scale (ESDS) to measure physical and social function, Jepsen, Schultz, Lusk, and McCorkle (1997) found that ESDS scores contributed significantly to the model of cancer survival. The ESDS scores also have been
linked with lower dietary intake in the lung cancer population (Sarna, Lindsey, Dean, Brecht, & McCorkle, 1993).

Adaptation Level and Nutritional Health
The adaptation level of a human adaptive system shows how well the system adjusts within its environment. The combined effects of the stimuli from the internal and external environments consist of the person’s adaptation level, and the adaptation level depends on the demands of the situation and the person’s current internal condition (Roy & Andrews, 1999). Studies have shown that nutritional health is linked strongly with morbidity and mortality (Liu, Bopp, Roberson, & Sullivan, 2002). Thus, nutritional health is conceptualized as an important aspect of the adaptation level indicating the human’s health and integrity.

Methods
Sample
This study was designed as a cross-sectional population-based study. The study protocol was approved by the university institutional review board. Participants were recruited from an inner-city senior housing complex in southern Connecticut. The facility supported independent living, and approximately one third of the residents were subsidized by public funds. Each resident in the facility was invited to participate. Residents were excluded if they were unable to give consent (n = 5), if they were hospitalized or too ill (n = 7), or if they were not English speaking (n = 26). Consent by proxy was not used because of its ethical and methodological difficulties.

At the end of a 7-month recruitment period, 243 respondents from an eligible population of 268 completed a structured in-home assessment by a trained geriatric nurse practitioner. To alleviate the respondent burden, the participants were given the option of breaking the interview into two sessions, and 2 of the 243 participants used this option. The participation rate was 91%. The nonparticipants (n = 25) were not significantly different from the participants in age (t = –.89, p = .38), gender (df = 1, p = .23), marital status (df = 3, p = .10), or ethnicity (df = 5, p = .74).

Instruments
Data were collected with a series of standardized measures (see Figure 1). Demographics collected in the interview included age, gender, marital status, ethnicity, education, living status, income levels, and religion. The Comorbidity Checklist was used to assess the presence of 14 chronic illnesses: myocardial infarction, angina, heart failure, other heart disease, hypertension, diabetes, arthritis, stroke, lung disease, vision problems, hearing problems, Parkinson’s disease, hip fracture, and cancer (Guralnik, 1989). Polypharmacy, operationally defined as the number of medications taken, was assessed by a medication review. The participants were asked whether, currently, they had taken any prescriptive or over-the-counter medications and, if so, to show the interviewer all these medications. The total number of medications was treated as a continuous variable.

Functional status was measured by the 10-item ESDS. The ESDS measures physical and social competence. Physical competence includes six activities: eating, dressing, walking, traveling, bathing, and toileting. Social competence includes home, work, recreational activities, and communication. The summed scores range from 10 to 51, with higher scores reflecting greater dependency. Reliability estimates (Cronbach’s alpha, .78–.96; test–retest r = .62) as well as established content, discriminant, and construct validity have been reported (Benoliel, McCorkle, & Young, 1980). Oral health was measured by the 12-item Geriatric Oral Health Assessment Index (GOHAI). The GOHAI assesses the dimensions of function (eating and speaking), pain, discomfort, worry, and social functioning. The summed scores on the GOHAI range from 12 to 36, with a high value indicating better self-perceived oral health. Reliability estimates (Cronbach alpha, .83–.79; interrater reliability alpha, .61) and established construct validity have been reported (Atchison & Dolan, 1990).

Satisfaction with support was measured by the subscale of the Social Support Questionnaire—Short Form (SSQ-SF). In six common situations, the participants were asked to list up to nine people who could be counted on (number score), and to specify their overall degree of satisfaction (satisfaction score). The coefficient alpha ranged from .90 to .93 for both number and satisfaction, and established construct and content validity have been reported (Sarason, Sarason, Shearin, & Pierce, 1987). The summed satisfaction scores ranged from 6 to 36, with a higher score indicating a higher degree of satisfaction with support (Sarason, Levine, Basham, & Sarason, 1983).

Depressive symptoms were measured by the 30-item Geriatric Depression Scale. Using a cutoff score of 11 or higher, the scale is 84% sensitive and 95% specific for diagnosing depression in the elderly (Yesavage, Brink, & Rose, 1983).

Nutritional status was measured by the 18-item Mini-Nutritional Assessment (MNA). The MNA contains a substantial component of anthropometric measurements as well as subscales for dietary behavior, general assessment, and subjective health. The summed score can be used as either a categorical or an interval variable. The scores, ranging from 0 to 30, categorize elders as well-nourished (≥24), at risk (23.5–17), or malnourished (<17). A sensitivity of 96%, a specificity of 98%, and satisfactory interrater reliability (Kappa coefficient, .65–.42) have been reported (Guigoz, Vellas, & Garry, 1996). Additionally, the percentage of age-weight loss in the preceding 6 months was calculated in the current study using a detailed weight history report.

Data Analysis
Data were analyzed using SAS, version 8 (SAS Institute, Cary, NC). Descriptive analysis and the PROC CALIS were performed. The models tested were covariance structure models with multiple indicators for all latent constructs. Means, standard deviations, ranges, and correlations for the study’s nine continuous manifest variables (e.g., age, number of medications and chronic illnesses as well as scores for oral health, depressive symptoms, functional status, and satisfaction with social support) are presented in Table 1.
The analysis followed a two-step procedure based on an approach recommended by Anderson and Gerbing (1988). In the first step of structural equation modeling (SEM), confirmatory factor analysis, was used to develop a measurement model. In the second step, the measurement model was modified to test the theoretical (causal) model of interest. For both steps, models were modified on the basis of the theoretical point of view and the path modification indices in the SEM.

**Results**

The sample was predominantly female, well-educated, and of diverse income and ethnic backgrounds. The ages ranged from 43 to 98 years (mean, 81.8 ± 9.4 years), and 44% of the participants (n = 107) were 85 years of age or older, with 50 participants (20.6%) in their 90's. Almost three fourths of the participants had earned at least a high school diploma. A description of the sample characteristics is presented in Table 2.

The results showed that the MNA scores ranged from 13.5 to 30 (mean, 24.1 ± 3.4). As classified by the MNA, 15 of the participants (6.2%) were malnourished, and 92 (37.9%) were at risk for malnutrition. In terms of weight loss, 131 (53.9%) of the participants reported that they had lost weight in the preceding 6 months, with 45 (18.5%) reporting weight loss equal to or exceeding 5% in the preceding 6 months. After descriptive analyses were completed, the SEM was performed by a two-step approach.

**Measurement Model**

The measurement model tested in this study consisted of three latent constructs: environmental stimuli, biopsychosocial responses, and adaptation level. The manifest indicators measuring the environmental stimuli were age (used as a continuous variable), education, ethnicity (non-white race), and number of chronic illnesses. The scores on the GOHAI, Geriatric Depression Scale, ESDS, and SSQ-SF subscale (measuring satisfaction with support), and the number of medications were used to measure the construct of biopsychosocial responses. Nutritional health, MNA scores, and the percentage of weight loss in the preceding 6 months were used to measure the construct in terms of the adaptation level.

All the indicators were significantly loaded on their corresponding latent constructs (p < .05) except the number of chronic illnesses (loading, .08). Although the number of medications was loaded relatively well on the construct of biopsychosocial responses (factor loading, 0.37), it was discovered that two variables (number of chronic illnesses and number of medications) showed significant inter-item correlations. Reviewing the proposition of the RAM, it became clear that biopsychosocial responses were the internal reactions of the human system to the outside stimuli. Because the number of medications and the number of chronic illnesses both are external, it was hypothesized that these two variables actually represent a different aspect of environmental stimuli. Thus, a decision was made to create a new construct, environmental stimuli 2 (F4), measured by the number of chronic illnesses and medications.

After model modification, the goodness-of-fit indices for the revised measurement model were 0.92 on the Non-Normed-Fit Index, 0.94 on the Comparative Fit Index, and 0.05 on Root Mean Square Error of Approximation (RMSEA), representing a good fit of the model to the data (Bentler, 1989). The small χ² value (relative to the degree of freedom) for the revised model was desirable: χ² (50, n = 243) = 87. The added construct F4 (environmental stimuli 2) actually performed well, with both indicators (number of chronic illnesses and medications) loading significantly on F4. The higher t value of F4, as compared with F3, suggested that F4 had a higher impact on the

<table>
<thead>
<tr>
<th>Variables</th>
<th>Age</th>
<th>NS</th>
<th>Loss</th>
<th>FS</th>
<th>DS</th>
<th>SS</th>
<th>O/R</th>
<th>Morb</th>
<th>Med</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nutrition</td>
<td>-.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Loss</td>
<td>.03</td>
<td>-.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Function</td>
<td>.29</td>
<td>-.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>.05</td>
<td>-.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Support</td>
<td>-.14</td>
<td>.36</td>
<td>-.14</td>
<td></td>
<td></td>
<td></td>
<td>-.24</td>
<td>-.48</td>
<td></td>
</tr>
<tr>
<td>Oral Health</td>
<td>-.03</td>
<td>.25</td>
<td>-.12</td>
<td></td>
<td></td>
<td></td>
<td>-.17</td>
<td>-.30</td>
<td>.18</td>
</tr>
<tr>
<td>Morbidity</td>
<td>.11</td>
<td>-.14</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.12</td>
<td>-.11</td>
</tr>
<tr>
<td>Number of Med</td>
<td>-.01</td>
<td>-.17</td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
<td>-.14</td>
<td>.30</td>
<td>-.09</td>
</tr>
<tr>
<td>Mean</td>
<td>81.8</td>
<td>24.1</td>
<td></td>
<td>3.0</td>
<td></td>
<td>20.3</td>
<td>6.7</td>
<td>39.9</td>
<td>33.2</td>
</tr>
<tr>
<td>SD</td>
<td>9.4</td>
<td>3.4</td>
<td>4.0</td>
<td>6.5</td>
<td>5.1</td>
<td>4.0</td>
<td>3.4</td>
<td>1.8</td>
<td>3.2</td>
</tr>
<tr>
<td>Range</td>
<td>43–98</td>
<td>13.5–30</td>
<td>0–21.7</td>
<td>10–38</td>
<td>0–27</td>
<td>12–36</td>
<td>18–36</td>
<td>0–9</td>
<td>0–17</td>
</tr>
</tbody>
</table>

Note. NS = MNA nutritional scores; Loss = % of weight loss in last 6 months; FS = Enforced Social Dependency Scale (ESDS) functional scores; DS = Geriatric Depression Scale (GDS) scores; SS = Satisfaction with support, subscale of the Social Support Questionnaire—Short Form (SSQ-SF) scores; O/R = Self-report Geriatric Oral Health Assessment Index (GOHAI) scores; Morb = number of comorbid conditions; Med = Number of prescriptive and nonprescriptive medications.
The revised measurement model was accepted as the study’s final measurement model.

**Theoretical Model**

The theoretical model tested in this phase replaced the double-headed arrows with single-headed, unidirectional arrows from construct to construct, indicating the prediction that the focal and contextual stimuli would have direct causal effects on the biopsychosocial responses, and that the biopsychosocial responses and focal stimuli would have direct causal effects on the adaptation level. The results indicated that the theoretical model demonstrated an acceptable fit with the data. Overall indices for the model were quite favorable, as indicated by the following values: a Goodness-of-Fit Index (GFI) of 0.94, an Adjusted Goodness-of-Fit Index (AGFI) of 0.91, a Comparative Fit Index of 0.94, a Non-Normed-Fit Index of 0.92, and a Root Mean Square Error of Approximation of 0.05. The \( \chi^2 \) statistic was desirable: \( \chi^2 (50, n = 243) = 88 \). The path coefficients all were statistically significant (\( p < .05 \)). The \( R^2 \) values showed that focal and contextual stimuli accounted for 18% of the variance in biopsychosocial responses, whereas biopsychosocial responses accounted for 35% of the variance in adaptation level (Figure 2).

The distribution of normalized residuals for the final model was quite symmetric and centered on zero. The parsimonious normed-fit index, reflecting both the parsimony and the fit of the final model, was 0.66. The relative normed-fit index was 0.99, which indicated that hypothesized causal relations between the structural variables provided an outstanding fit to the data. Consequently, this theoretical model was accepted as the study’s final model (see Figure 2).

**Discussion**

The specific aim of this study was to examine the goodness-of-fit exhibited by the proposed RAM-based framework of nutritional health for community-dwelling elders. Through the use of SEM analytical techniques, both the links among the latent constructs and the relations among all the manifest indicators were tested. Conceptually, the resulting model demonstrated that the focal and contextual stimuli influenced biopsychosocial responses and subsequently had an impact on the adaptation level. Theoretically, the results indicated that functional status, oral health, depressive symptoms, and satisfaction with support affected nutritional health directly. Oral health, depressive symptoms, functional status, and satisfaction with support mediated the effects of age, ethnicity, education, and the number of medications and chronic illnesses on nutritional health. The model accounted for 35% of the variance in nutritional health and demonstrated a good fit with the data. In terms of theory testing, two propositions of the RAM were supported: (a) four biopsychosocial response modes are interrelated, and (b) adaptation level is the function of environmental stimuli and biopsychosocial responses.

**Interrelated Biopsychosocial Response Modes**

Assessment of biopsychosocial responses indicated how an individual manages to cope with environmental stimuli. Responses in one mode may have an effect or act as a

<table>
<thead>
<tr>
<th>TABLE 2. Sample Demographic Characteristics (N = 243)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong></td>
</tr>
<tr>
<td>Age group</td>
</tr>
<tr>
<td>Less than 65 years</td>
</tr>
<tr>
<td>65–74 years</td>
</tr>
<tr>
<td>75–84 years</td>
</tr>
<tr>
<td>85 years and above</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Ethnicity</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Black</td>
</tr>
<tr>
<td>Hispanic</td>
</tr>
<tr>
<td>Asian</td>
</tr>
<tr>
<td>Native American</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Marital status</td>
</tr>
<tr>
<td>Widowed</td>
</tr>
<tr>
<td>Married</td>
</tr>
<tr>
<td>Divorced</td>
</tr>
<tr>
<td>Single</td>
</tr>
<tr>
<td>Education (years)</td>
</tr>
<tr>
<td>8 or less years</td>
</tr>
<tr>
<td>9–11 years</td>
</tr>
<tr>
<td>12 years</td>
</tr>
<tr>
<td>13–15 years</td>
</tr>
<tr>
<td>16 years or more</td>
</tr>
<tr>
<td>Living arrangement</td>
</tr>
<tr>
<td>Alone</td>
</tr>
<tr>
<td>With others</td>
</tr>
<tr>
<td>Annual income (US $)</td>
</tr>
<tr>
<td>Less than 10,000</td>
</tr>
<tr>
<td>10,000–20,000</td>
</tr>
<tr>
<td>20,001 or more</td>
</tr>
<tr>
<td>No idea or prefer not to answer</td>
</tr>
<tr>
<td>Religion</td>
</tr>
<tr>
<td>Catholic</td>
</tr>
<tr>
<td>Protestant</td>
</tr>
<tr>
<td>Jewish</td>
</tr>
<tr>
<td>Buddhist</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

biopsychosocial responses (F2) and the adaptation level (F1). The two constructs were relabeled to reflect their contributions to the outcome of interest. To be in line with the propositions of RAM, the F3 was named as contextual stimuli and the F4 was named as the focal stimuli. This
stimulus for one or all of the other modes. The RAM asserted that four modes (physiologic, self-concept, independence, and role function) are interrelated (Roy & Andrews, 1999). In the current study, the manifest indicators representing these different modes were statistically correlated, and all loaded significantly on the construct of biopsychosocial responses. It also was significant clinically that oral health, depressive symptoms, functional status, and satisfaction with support had a joint influence on nutritional health.

Impact of Environmental Stimuli and Biopsychosocial Responses

The proposition concerning the impact of environmental stimuli and biopsychosocial responses was tested in this model. The focal and contextual stimuli provided a reasonable explanation for the biopsychosocial responses ($R^2 = .18$). The unknown factors that contributed to the rest of the variance in biopsychosocial responses were, in the words of RAM, “residual stimuli.” Several potential factors were identified and collected including gender, living status, and income levels. However, the categorical nature of these variables and limited variability in gender (80% female) and living status (92% living alone) made it impossible for the final SEM model to be tested in terms of their effects on nutritional health. Future analysis would be improved if such variables were more evenly distributed.

The construct that had the most impact on biopsychosocial responses was labeled as focal stimuli. Because F4 had a higher path coefficient than F3, F4 was labeled as “focal stimuli.” Contextual stimuli were the other stimuli in the situation that contributed to the effect of the focal stimuli or influenced how the human system dealt with the focal stimuli (Roy & Andrews, 1999). In that sense, the resulting model, with age, ethnicity, education, and non-white race labeled as contextual stimuli, was noteworthy. The importance of these demographic variables for nutritional health has been stressed in the literature. Initially, variables such as age were postulated as the focal stimulus given their acceptance as a major determinant. The data, however, failed to support the long-lasting assumption that age has a detrimental effect on nutritional health. This finding is important because age, education, and ethnicity are not reversible. Conversely, the numbers of chronic illnesses and medications, emerging as the focal stimuli, are potentially modifiable by better preventive care and improved clinical practices. Additional research is needed for an understanding of these relations and the ramifications of the findings.

The adaptation level is a combined effect of the environmental stimuli and biopsychosocial responses (Roy & Andrews, 1999). The model empirically supported this proposition, suggesting that the environmental stimuli had an impact on the biopsychosocial responses and

![Figure 2](image-url)
subsequently on the adaptation level. Biopsychosocial responses acted as a mediator influencing the adaptation level. Depressive symptoms, satisfaction with support, functional status, and oral health emerged as significant factors for nutritional health.

Effective treatment options exist for elders with depressive symptoms once they are identified. Routine screening for depressive symptoms should be standardized. There also is a growing recognition of the importance of social support to health. Referral for counseling services, social activity programs, and community outreach resources may be beneficial. Interventions should be tailored for different types of social isolation, and arrangements should be made to remove barriers such as transportation.

Functional status has demonstrated its power to predict nutritional health. Each member of the interdisciplinary geriatric team could be extremely helpful in assessing, maintaining, and maximizing the functional status of community-dwelling elders (Chen, Kenefick, Tang, & McCorkle, 2004). Prior research has demonstrated the effectiveness of such collaboration in improving function, retaining elders at home, and enhancing other related patient outcomes (Stuck, Egger, Hammer, Minder, & Beck, 2002). Oral health also has played an important role in relation to health. Recent findings show an association between cardiovascular disease and oral health, which may be related to the insufficient produce intake and altered diet pattern of edentulous subjects. Participants with poor dental status should be given instructions for maintaining dietary quality (e.g., shredding or blending of fruits and vegetables) (Ritchie, 2002).

Despite its careful design, this study had certain limitations. The sample was confined to one data collection site, with the result that 56% of the participants were Jewish. Therefore, the generalizability of the study results is limited to these elders. Furthermore, although in the SEM, all the paths were marked with arrows, implying causal relations, it should be noted that given the cross-sectional nature of this study, these relations were correlational rather than causal. Future studies would be improved by using longitudinal data, and by including a boarder range of demographic variations in the study sample. Nevertheless, the sample size of 243 in a post hoc analysis showed that the power was .86. This level of power enhanced confidence for making inferences from the study results. Many factors were studied and analyzed simultaneously, which provided a unique opportunity for investigating complicated relations among these factors.

This study was designed as a theory-guided and population-based investigation to evaluate the multifactorial etiology of nutritional health among community-dwelling elders. The findings provide tangible evidence that the etiology of nutritional health is multifactorial, and that it encompasses physiologic (e.g., chronic illness, medications taken, oral health, functional status), social (e.g., satisfaction with support), psychological (e.g., depressive symptoms), and cultural (e.g., ethnicity, education) elements. Emphasis should be placed on the fact that the study sample was well nourished (55.9%), indicating that the factors found in this study also contributed to a positive nutritional health aspect of the adaptation level. Therefore, an understanding of nutritional health etiology will assist nurses in developing interventions for preventing, treating, and minimizing nutritional problems and related disabili-

One of the most significant proposition statements of RAM is that “the goal of nursing is the promotion of adaptation in each of the four modes, thereby contributing to the person’s health, quality of life, and dying with dignity” (Andrews & Roy, 1999, p.20). This framework of nutritional health among community-dwelling elders provides a theoretical and empirical roadmap for such efforts. More nurses without community health backgrounds will be caring for elders in the community, and frameworks such as this are needed to guide practice.

There is a pressing need for further research on nutritional health among community-dwelling elders. Key issues have not been formally studied, and previous research does not provide definitive answers. Future studies should replicate and test the application of this model over time to develop nutritional interventions that in turn promote adaptation, independence, and quality of life for community-dwelling elders.

References


Evaluating In-Home Training for Fathers of Children With Autism Using Single-Subject Experimentation and Group Analysis Methods

Jennifer H. Elder • Gregory Valcante • Hossein Yarandi
Deborah White • Timothy H. Elder

**Background:** Autism, or the broader category of autistic spectrum disorder, is a complex developmental disability with uncertain etiologies that appears to be increasing in prevalence. Researchers have stated that training programs for children with autism are most effective when they are individualized, address communicative intent of child behaviors, and promote social reciprocity between children and individuals with whom they have regular contact. Yet, to date, most of what is known comes solely from studying mothers, who have traditionally been the most accessible parent.

**Objectives:** In this study the mother–child in-home training program was modified and evaluated for its effects on the acquisition of training skills by fathers and on precommunication skills by the autistic children.

**Methods:** Frequency counts of skills taught to fathers and targeted child behaviors were obtained from videotaped father–child play sessions. These data were analyzed for each father–child dyad by using graphs and visual analyses, which are integral parts of single-participant experimentation. This procedure was replicated across all of the father–child dyads. Data were then grouped and analyzed using the more traditional repeated measures analysis of variance.

**Results:** The most significant findings were increases in father use of imitating with animation \( p < .0001 \) and child initiating following training \( p < .0004 \). Also noteworthy were significant increases in father responding \( p < .0005 \) and child vocalizations \( p < .05 \).

**Discussion:** Results of the study indicate that the in-home training for fathers of children with autism was effective and valued by the participating families.

**Key Words:** autism • father–child interactions • parent training • single-subject experimentation

**Theoretical Framework and Key Concepts**

Defining **Social Reciprocity and Its Theoretical Origins**

Social reciprocity, a term often associated with social interaction and attachment theories, describes the bidirectionality...
of adult–infant interactions. Calhoun, Rose, Hanft, and Sturky (1991) defined it as the ongoing interaction process between infants and their primary caregivers. Various aspects of social reciprocity, including joint attention between the caregiver and child, have been examined over the years by a number of researchers (Barnard, 1983; Brazelton, Koslowski, & Main, 1974; Mundy, Sigman, & Kasari, 1994). Maternal responsiveness has been shown to correlate overwhelmingly with positive child outcomes (e.g., cognitive and social competence, verbalization, and cognitive abilities).

Social Reciprocity (Turn-Taking) Training Component
Parent–child turn-taking, a component of social reciprocity, is described in classic works as an essential factor in child language development (Bruner, 1973; Snow, 1983) and appears to be particularly important for autistic children, who typically exhibit severe language delays. It has been hypothesized that parents of autistic children may not engage in turn-taking procedures with their children because the autistic child does not respond (take his or her turn) as expected and the parent does not receive reinforcement to continue the interaction (Cunningham, Reuler, Blackwell, & Deck, 1981).

Sigman, Mundy, Sherman, and Ungerer (1986) reported that longitudinal studies suggest that the capacity for joint attention, a core construct, is linked to language acquisition. This finding is consistent with social interaction theory and has important implications for families who have ongoing contact with their autistic children in familiar home environments. If joint attention and social reciprocity are essential precursors to language acquisition in children with autism and language acquisition is an important indicator of long-term prognosis and functionality, more research is needed to evaluate family-conducted interventions that specifically target facilitation of social attending, initiating, and responding. Figure 1 illustrates how specific training components are theoretically linked.

Parent-Training Interventions for Children With Autism
Parent training (PT) has been empirically validated as a cost-effective treatment technique for teaching parents to modify problematic child behaviors and facilitate child learning (Forehand & McMahon, 1981; Patterson, 1982). The growing popularity of PT programs during the 1990s resulted in numerous empirical studies suggesting that collaborative approaches where parents are active participants at all stages of the treatment process are more effective than traditional didactic teaching and family therapy models (Webster-Stratton & Herbert, 1993).

Most in-home PT interventions for children with autism comprise one of three main categories (Elder, 2002). The first comes from behavioral psychology where operant conditioning has been used and has involved replications of Lovaas’ (1978) intensive (30–40 hr/week) discrete trial (ABAB) training. The second, also from behavioral psychology, used a functional analysis approach to target individual child behaviors (Blakeslee, Sugai, & Gruba, 1994; Gable, 1996; Vollmer & Northup, 1996; Wacker et al., 1992). The third category addressed gaps in developmental processes, emphasized language development, used interventions that specifically target developmental deficiencies, and taught parents to use normal activities to teach children in their natural environments (El-Ghoroury & Romanczyk, 1999; McConnell, 2002). Of the approaches mentioned, replications and home adaptations of Lovaas’ approaches (first developed in the 1970s and initially implemented solely by professionals) are most plentiful. However, it is difficult for parents to implement such time-intensive programs and comparisons of these parent-directed adaptations and professional-directed interventions have shown that professionally implemented interventions are usually more effective (Bibby, Eikeseth, Martin, Mudford, & Reeves, 2001; Smith, Buch, & Gamby, 2000). Also, whereas improvements in child intelligence scores have been reported in both professional-directed and parent-directed groups, little is known about long-term maintenance effects of acquired skills and results on long-term child social development. Similar issues are raised by functional analysis methods that may target specific behaviors important for individuals but not address broader developmental processes or provide information on long-term developmental effects.

Of the three types of approaches, the third is more general, builds on well-established developmental concepts and theories, and is less burdensome for parents and children because it uses low demand learning situations such as child-directed play sessions and/or socialization outings. While normal child development literature provides much empirical evidence regarding various aspects of child social and language development, it is less clear if these theories fully apply to children with autism. Preliminary evidence

**FIGURE 1.** Conceptual model illustrating linkages between theoretical framework and specific training components.
has shown that interventions targeting pivotal skills such as joint attention may produce positive sustained long-term developmental effects in autism (Mundy et al., 1994). Further, it may be possible to train parents to effectively teach these important skills (Koegel, Koegel, & Dunlap, 1999; McConnell, 2002) during everyday interactions.

The Role of Fathers in Training Autistic Children

Most of what is known about the effects of PT programs in autism comes from studying mothers, who have traditionally been the primary caretakers and the most accessible parent. Over the past 20 years, however, there has been mounting interest in the role of fathers and the effect of their involvement on child development generally (Lamb, 1987; Elder, Valcante, Groce, Yarandi, & Carlton, 2002). Tiedge and Darling-Fisher (1996) reported that healthy father–child interactions positively affect development. However, Booth and Crouter (1998) and Damon (1998) noted in their reviews of contemporary literature that, whereas some studies support this positive view, other research shows no clearly discernable paternal effects on child development. These conflicting findings are further complicated in the case of fathers of autistic children about whom even less is known.

Currently, literature emphasizes the role of mothers in child training and, as mentioned above, there is a dearth of studies on fathers. Furthermore, fathers frequently assume secondary child-training roles, relinquishing primary training responsibility to the mothers. Yet, it is likely that fathers have the potential to positively influence the immediate and long-term welfare of their autistic children. The primary purpose of this research was to evaluate the effects of an in-home training program on the acquisition of training skills by fathers of autistic children and of precommunication skills by the children themselves.

Methods

Participants and Setting

Fathers of 18 children with autism were recruited through the Centers for Autism and Related Disabilities and/or a community family health practice. Parental informed consent and child assent were obtained for each family. Child eligibility was assessed using medical records, in-depth intake evaluations conducted by a multidisciplinary team composed of a child psychiatric nurse, pediatric nurse practitioner, child psychiatrist, and special education teacher and parent interviews with the principal investigator and coprincipal investigator. Child inclusion criteria included (a) a diagnosis of autistic disorder according to Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) criteria (APA, 1994), and above standard established cutoff scores on the Childhood Autism Rating Scale (CARS) (Schopler, Reichler, DeVillis, & Daly, 1980), and Autism Diagnostic Interview—Revised (ADI-R) (Lord et al., 1994); and (b) residence within 150 miles of the university. Children were excluded if their medical histories and/or physical examinations indicated that they had physical or sensory impairments or significant medical problems (e.g., seizure disorders, chronic otitis media).

Fathers were defined as biological fathers residing with the autistic child or male caregivers who had resided continuously with the autistic child for at least 6 months. Exclusion criteria included any reported physical, major psychiatric, or sensory problems (e.g., speech and language disorders, hearing loss) that might affect the ability to conduct training and/or interact with their children.

Training and data collection occurred in the room in the participants’ homes where the children were most often exposed to informal family interactions (e.g., family room, den). Only the autistic child and father were present during the sessions. The room was set up to facilitate videotaping. A standard rolling duffel bag of toys was used for each session. Toys were selected for inclusion based on recommendations of experienced child psychiatry inpatient staff as well as parents of children with autism who participated in a previous study (Elder, 1995).

Instruments

In addition to collecting sociodemographic information for each family, two well-established instruments were used to describe autistic features of the child participants.

Childhood Autism Rating Scale The CARS is an instrument used to assess autism, consisting of 15 items and a 7-point Likert-type scale. Each item contributes equally to the total score. The CARS items cover the following behaviors related to one of the following areas: relationships with others; imitation, emotional expression, body use, pecularities in object use; resistance to change; visual, auditory, and tactile responsiveness; anxiety; verbal and nonverbal communication; activity level; and intellectual ability. Possible scores range from 15 to 60. As recommended by the instrument developers, children with scores more than 30 were classified as autistic. Instrument psychometrics are acceptable and well documented (Schopler et al., 1980). The CARS manual reports a test–retest reliability of .88 on total CARS scored based on 91 cases over a 12-month period. The alpha coefficient for the total score is reported as .94. Validity has been assessed under various conditions and suggests that CARS is appropriate for use as a widespread screening device (Schopler et al.).

Autism Diagnostic Interview—Revised The ADI-R (Lord et al., 1994) is a standardized parent interview for assessing the presence and severity of symptoms. The validity of the instrument is based on the diagnostic criteria for autism in the DSM-IV (APA, 1994) and ICD-10 symptom domains. The ADI-R covers impairment in reciprocal social interaction, communication, repetitive behaviors, and stereotyped patterns. A diagnosis of autism is established if an individual scores at or above the cutoff score in the three ICD-10 symptom domains. Lord et al. reported multivariate weighted kappa levels of reliability from .62 to .89. The Cronbach alpha measure of internal consistency yielded the item–total correlation for social area range from .54 (direct gaze) to .77 (quality of social overtures) with an alpha level of .95. The item—total correlations for restricted and repetitive behaviors range from .30 to .53.
with the reported alpha level of .69. The item correlations ranged from −.06 to .77 with an alpha level of .85. Communication items for all participants showed item—total correlation ranging from .45 to .70 with an alpha level of .84. The PI or a consulting child psychiatrist administered the ADI-R. Both were trained by, and established reliability with, developers of the instrument.

Dependent Variable Measurement

Behavioral response categories from a previous mother–child study (Elder, 1995) were combined to provide a more comprehensive constellation of behaviors. These revised operational definitions were used in a study comparing child and parent social reciprocity behaviors in clinic setting versus home setting (Elder, Valcante, Groce, Yarandi, & Carlton, 2002). The definitions appear in Table 1.

Therapy Attitude Inventory (TAI) assesses parental satisfaction with the process and outcome of parent training. Specific areas assessed are quality of parent–child relationship, the child's behavior, and overall family adjustment. Responses to the 7-item survey are based on a 5-point Likert scale with 0 indicating no satisfaction (no improvement) and 5 indicating high satisfaction (great improvement). The Cronbach alpha was calculated at .88 by Eyberg and Robinson (1982) and Eyberg and Boggs (1989). Families completed this instrument at the end of the in-home training program.

Independent Variable (Father-Training Intervention)

The father-training intervention was taught by the first author and contained two components: “imitating/animating” (I/A) and “expectant waiting” (E/W).

Imitating/animating was operationally defined as “The movement cycle that begins within 5 seconds of a child's initiation (e.g., child utters a vocalization, jumps up and down) wherein the father imitates the child's behavior in an animated manner (e.g., exaggerated affect, lively movement).” In earlier research and clinical work, parents described not knowing how to play with their children as they sat passively in the play sessions or aggressively tried to direct interactions, not allowing their children sufficient time to respond (Elder, 1995). Teaching I/A as the first intervention addressed this concern by promoting basic turn-taking play interactions that were enjoyable and helped parents relax in the play sessions. Play sessions were then videotaped with booster teaching sessions incorporated when fathers demonstrated low frequencies or difficulties using the skills taught. The PI then taught Intervention 2, “expectant waiting,” to the father. Expectant waiting was defined as “The movement cycle that begins when the father prompts a child behavior (e.g., ‘Say ball’; ‘Find block.’) and waits at least 3 seconds—during which time the parent provides the child with facial positions signaling the availability of positive social interactions (e.g., attention focused on the child, eager parental facial expression)—and ends with the desired child response or a parental prompt (e.g., repeating the prompt, labeling the object) if the response does not occur.” Again, a series of play sessions was recorded with booster teaching sessions provided. The booster sessions included a review of graphed data from previous sessions and videotapes of the fathers using the skills taught by the PI with their own children. By viewing videotapes, fathers were able to see what they were doing well and what needed improvement. (Figure 2 illustrates the training sequence.)

Research Design: Description of, and Rationale for, Using a Combination of Single-Subject (Participant) Experimentation and Group Analysis

The research design chosen for this study included both single-subject experimentation methods to evaluate intervention component effects in individual participants and a more traditional group analysis (i.e., RANOVA) to assess external validity of the intervention across all father–child dyads. The key element of single-subject experimental design is the introduction and manipulation of an independent variable under controlled conditions (Cook & Campbell, 1979). This type of experimentation has a long tradition in applied and clinical behavioral research and is indicated when the research question addresses the behavior of individuals rather than group behavior, individual characteristics of participants vary greatly, and/or representative sampling of individual behaviors is difficult to achieve (Elder, 1997; Hersen & Barlow, 1987; Tawney & Gast, 1984).

<table>
<thead>
<tr>
<th>Behavior</th>
<th>Operational Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Father initiating</td>
<td>The movement cycle that begins with the father orienting toward, reaching for, talking to, or offering items to the child, and ends when the child has an opportunity to respond to the parent's initiation</td>
</tr>
<tr>
<td>Father responding</td>
<td>The movement cycle that begins with the occurrence of a child behavior and ends after the father initiates the requested movement or offers a comment that indicates that the parent acknowledges, understands, and/or approves of the child initiation or behavior within 2 s</td>
</tr>
<tr>
<td>Child initiating</td>
<td>The movement cycle that begins with a child orienting toward, reaching for, talking to, or offering items to the father, and ends when the parent has an opportunity to respond to the child's initiation</td>
</tr>
<tr>
<td>Child responding</td>
<td>The movement cycle that begins with the occurrence of a father prompt and ends after the child initiates the requested behavior within 5 s</td>
</tr>
<tr>
<td>Child vocalizations</td>
<td>Any audible vocal sounds emitted by the child excluding laughing, crying, coughing, yawning, or sneezing, and bounded by a father vocal response or a 2-s pause</td>
</tr>
</tbody>
</table>
This approach uses the participant's behavioral baseline as his or her own control. Thus, instead of comparing one participant's pre–post difference to that of another, as in the control group design, the key comparison in single-participant experimental design is between the participant's baseline and treatment measures. Therefore, issues related to participant selection and matching are not particularly relevant. Another advantage of this within-subject approach is that every participant receives intervention (i.e., there are no untreated controls, just participants’ baselines on which intervention has not yet begun) (Iwata, Dorsey, Slifer, Bauman, & Richman, 1982).

Johnston and Pennypacker (1993) assert that because behavior occurs at the level of the individual, researchers of human behavior must have as their goal the understanding of the individual's interaction with his or her environment. Single-subject experimentation affords this type of close inspection of individual participants. Once that is accomplished, the question of generalizing the results from one or a few individuals to a larger group can be addressed by replicating across participants (N = 18).

There are several reasons why a specific type of single-subject experimentation, the multiple-baseline (MBL), was used to assess the efficacy of the two training components, I/A and E/W. First, it is not necessary to reverse or withhold an intervention to demonstrate experimental control (Elder, 1997; Hersen & Barlow, 1987; Tawney & Gast, 1984). Furthermore, behaviors such as learned skills are functionally irreversible (cannot be “unlearned” and cannot be withdrawn). Also, because MBL involves repeated measures of behavior both prior to and following intervention, the design allows one to establish a stable level of performance, thus ruling out sampling errors, practice effects, habituation, and other possible confounds that may make one or two measures suspect or unrepresentative of the phenomenon of interest. The staggered (delayed) introduction of intervention across the baselines provides a compelling argument against the influence of confounding variables, especially given the fact that measures continue to be taken throughout. This method is particularly useful when evaluating specific components of new intervention packages, allowing investigators to make changes when indicated and provide feedback to participants without compromising the scientific integrity of the experiment (Hersen & Barlow, 1987; Tawney & Gast, 1984).

Training and Data-Collection Sequence

The training and data-collection sequence are depicted in Figure 2. After informed consent and assent was obtained and the initial clinic evaluation was performed, a series of baseline sessions was filmed in the homes with each father and child under the same conditions as the subsequent postintervention sessions. The number of baseline sessions ranged from 4 to 8 to ensure that data trends were stable prior to beginning the intervention. As previously discussed, after baseline, the PI taught Intervention 1, “imitating with animation,” to the father. Mothers were not present during the father training. After training, fathers were instructed to use this strategy in everyday interactions with their autistic children. The same training and data-collection method was used for Intervention 2, “expectant waiting,” which was taught after “imitating with animation.”

Research assistants, who were extensively trained in videotaping techniques, camera placement, and maintaining unobtrusiveness, videotaped each child with the father for 15 min during each home visit. The total number of videotaped sessions for each of the 18 father–child dyads ranged from 14 to 22 over a 10- to 12-week period. A summary of the training and videotaping sequence is shown in Figure 2. Ten-minute samples of each 15-min session were coded. The first 2-min segment was considered a settling-in period and was not coded, whereas the last 3 min were used to facilitate agreement of independent raters, thus preserving scientific integrity of the data-collection method. Research assistants recorded occurrences of father and child reciprocity behaviors onto computer-generated forms by using specific behavioral definitions and response classes. Ratings were conducted independently and behavioral frequency counts determined by the number of occurrences divided by the length of the session (10 min). Coding and analysis of target behaviors were facilitated using the Multi-Option Observation System for Experimental Studies (MOOSES) (Tapp, 1996) computer observation coding system. Using this software,
data were entered onto laptop computers, labeled and organized, and stored on a larger desktop computer with network access. The PI was aware of the sequence of tapes from each father and child, and the father–child progress was followed. This is customary in single-subject experimental methodology. To minimize the potential for bias, behavioral observers were blind to the training condition.

The behavioral observer training methods used in this study were previously developed, tested, and reported by Elder (1999). The training followed this sequence: (a) familiarization with the behavioral-response categories as defined in the coding manual, (b) practice coding videotapes of role-played interactions that clearly portray correct use of the father-training skills, and (c) practice coding preexisting parent–child videotapes not associated with the current project. The videotapes used in Steps b and c were previously coded by the PI and have well-established interrater reliability. The trainees were required to reach a criteria level of 90% (agreement with preestablished frequencies) on each of the tapes. To evaluate for possible rater drift throughout the course of the project, “drift checks” were conducted. This involved a second rater who coded 20% of the videotaped sessions that were randomly selected by the PI. When interrater agreement fell below 80%, practice sessions were reinstituted until the 90% criteria level was met again. The PI developed this training procedure by adapting the Eyberg and Robinson (1982) procedure and piloting trials with 12 novice observers (undergraduate nursing students) who were taught to observe and record a variety of child, parent, and staff behaviors at the University of Florida’s Children’s Mental Health Inpatient Unit (Elder, 1999). Results of interrater reliability “drift checks” yielded the following: father initiating = 89.9, father responding = 79.2, father I/A = 82.0, father E/W = 75.8, child initiating = 77.7, child responding = 77.8, and child vocalizations = 83.0. Overall interrater agreement was 80.2.

**Results**

There were 14 boys and 4 girls from varied ethnicities (Asian American, n = 3; Hispanic, n = 2; African American, n = 3; and European American, n = 10). The larger number of males reflects the fact that autism occurs more often in males than in females (Cohen & Volkmar, 1997). The children ranged in age from 24 to 84 months, M = 56.7 months (SD = 17.48 months). Fathers ranged in age from 24 to 49 years, M = 36.1 years (SD = 7.6 years). Family background were working class (n = 4), middle class (n = 12), and upper-middle class (n = 2). All but 2 of the 20 originally enrolled families completed the entire 10- to 12-week in-home training protocol. (Attrition of the two families was due in one case to a child requiring unexpected surgery and in the other case to the other family’s relocating.)

**Results Using Single-Participant Experimental Methodology**

Visual analysis is an integral part of single-subject experimentation methodology (Elder, 1997; Tawney & Gast, 1984) and, in this case, yielded important, clinically relevant findings that were immediately incorporated into in-home training. As customary in single-subject experimentation, videotapes and data were examined thoroughly on an ongoing basis throughout the course of the study. This provided an efficient mechanism for training fathers who could clearly see trends of their own behavior as well as child responses and make adjustments as necessary. Also, through the use of visual analyses, it was possible to see effects of the individual training sessions in each experimental condition. This information is not clearly ascertained when examining results of the grouped statistical analyses.

Figure 3 reflects and illustrates one father’s skill acquisition. This graph was selected for discussion because it illustrates findings commonly seen in other father–child dyads. Results show that the father did not use either strategy during baseline, a common finding among fathers. Frequency rates of I/A increased, also noted for this father, markedly from baseline while EW rates increased initially when first taught in Condition 2 but later decreased. Visual inspection helped illustrate why EW rate was not maintained in the last Condition 2 session, where there was some covariation between the increased uses of I/A and the decrease in EW rates.

Graphed data (Figure 4) and videotapes of the father–child sessions were discussed with the father during feedback sessions. The father stated that he enjoyed using I/A but that EW was difficult. This is likely due to the fact that this father had traditionally relied heavily on questioning and therefore found inserting EW into the interactions to be unnatural, especially when focusing on the implementation of I/A.

**Results of Replications Across Participants**

The second level of analysis involved group data across the 18 children and parents. Grouped RANOVA data are presented in Tables 2 and 3. Because there was variation in the number of sessions in each condition, individual sessions frequencies were aggregated and values for each condition were compared.

**Father Behaviors**  Significant increases in father I/A rates were noted from baseline to Condition 1 (p < .0001) and from baseline to Condition 2 (p < .0001). There were no significant changes in father EW rates from baseline to Condition 1 because EW was not taught to fathers until Condition 2. After training, there were significant increases in EW rates from baseline to Condition 2 (p < .006). With father initiating, there were no significant differences when comparing baseline versus Condition 1 and baseline versus Condition 2, whereas significant increases in father responding rates were noted when comparing baseline father responding with Condition 1 (p < .0005) and Condition 2 (p < .0021).

**Child Behaviors**  Child initiating increased significantly from baseline to Condition 1 (p < .0004). Significant increases were noted also from baseline to Condition 2. Child responding did not increase significantly from baseline to Condition 1 or 2. Increases were seen when comparing child vocalization frequencies in baseline versus Condition 1 (p < .045), but not when comparing baseline with Condition 2 (p < .058).
Social Validity of In-Home Training
Each family completed a TAI survey that was a measure of social validity. The medians for all items were equal to, or exceeded, 4.0, with 5.0 (the highest possible rating) noted as the median for the last item, which addresses overall parental satisfaction with the program.

Discussion
Findings indicate that fathers acquired and successfully implemented the training skills they were taught and that the autistic children exhibited improvement in their precommunication skills. The most significant findings were increases in father I/A rates (>500%), and child initiating (>100%). Because deficits in social initiating are considered to be a sign of autism, this finding suggests possible directions for future work.

Given the dramatic increases in child initiating, it was surprising to see no significant increases in child responding rates from baseline to Condition 1 or 2. This finding requires further investigation and may, at least in part, reflect the difficulty coders encountered in detecting child

FIGURE 3. A visual analysis of Father 1 data for imitating/animating (IA) and expectant waiting (EW) interventions.

FIGURE 4. Mean frequencies of statistically significant father and child behaviors across conditions (N = 18).
Also noteworthy are significant increases in father responding rates (60%) and child vocalization frequencies (>50%). The increases in father responding rates were expected because the intervention is designed to enhance parental awareness of the child’s initiations. The significant increases in child vocalization frequencies following implementation of I/A for fathers were also anticipated and consistent with results of Elder’s (1995) mother-training study.

Also important was the finding of no significant differences for father initiating frequencies when comparing baseline versus Condition 1 and baseline versus Condition 2. Because the interventions were designed to promote balanced parent–child interactions, this finding reflects the expectation that the frequency of father initiations would not increase across conditions. This also suggests that fathers may benefit from an intervention component that specifically teaches them to initiate less because frequent parental initiating may negatively affect parent–child social balance and limit opportunities for child initiating and vocalizations.

Social validity data indicate that parents viewed the interventions favorably and anecdotal reports attest to the intervention clinical significance. For example, during one in-home session, the autistic child made eye contact with the father and said “Daddy” for the first time in the child’s life. Another father related how after training he “automatically” began imitating his child with animation and modeled this for his wife as well as other families of children with autism in informal social settings. A third family

<table>
<thead>
<tr>
<th>Response Variables</th>
<th>Baseline</th>
<th>Condition 1</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Father initiating</td>
<td>83.6</td>
<td>81.6</td>
<td>0.57</td>
<td>.638</td>
</tr>
<tr>
<td></td>
<td>26.0</td>
<td>33.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>32.0</td>
<td>41.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>139.2</td>
<td>147.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Father responding</td>
<td>25.0</td>
<td>40.6</td>
<td>18.21</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>13.7</td>
<td>19.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>7.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>53.0</td>
<td>73.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Father imitating with animation</td>
<td>3.3</td>
<td>19.1</td>
<td>26.19</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>4.5</td>
<td>14.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19.0</td>
<td>54.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Father expectant waiting</td>
<td>2.7</td>
<td>3.0</td>
<td>0.15</td>
<td>.702</td>
</tr>
<tr>
<td></td>
<td>3.3</td>
<td>3.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.50</td>
<td>11.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child initiating</td>
<td>12.28</td>
<td>25.3</td>
<td>19.24</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>9.1</td>
<td>16.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>3.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>31.0</td>
<td>57.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child responding</td>
<td>29.5</td>
<td>29.3</td>
<td>1.07</td>
<td>.391</td>
</tr>
<tr>
<td></td>
<td>13.5</td>
<td>11.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>11.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>62.5</td>
<td>51.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child vocalizations</td>
<td>30.1</td>
<td>46.3</td>
<td>4.67</td>
<td>.045</td>
</tr>
<tr>
<td></td>
<td>17.8</td>
<td>31.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>8.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>80.8</td>
<td>134.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Statistically significant increase was detected between baseline and Condition 1. Fathers had not yet been taught expectant waiting.
TABLE 3. A Comparison of Baseline and Condition 2 Response Variable Frequencies for Father–Child Dyads (N = 18)

<table>
<thead>
<tr>
<th>Response</th>
<th>Variables</th>
<th>Baseline</th>
<th>Condition 2</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Father initiating</td>
<td>M</td>
<td>83.6</td>
<td>91.2</td>
<td>0.57</td>
<td>.638</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>26.0</td>
<td>29.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>32.0</td>
<td>41.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>139.2</td>
<td>133.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Father responding</td>
<td>M</td>
<td>25.0</td>
<td>41.8</td>
<td>13.18</td>
<td>.002</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>13.7</td>
<td>23.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>1.0</td>
<td>6.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>53.0</td>
<td>95.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Father imitating with animation</td>
<td>M</td>
<td>3.3</td>
<td>16.5</td>
<td>26.03</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>4.5</td>
<td>12.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>0.0</td>
<td>3.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>19.0</td>
<td>44.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Father expectant waiting</td>
<td>M</td>
<td>2.7</td>
<td>6.0</td>
<td>9.64</td>
<td>.006</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>3.3</td>
<td>3.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>9.5</td>
<td>14.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child initiating</td>
<td>M</td>
<td>12.3</td>
<td>26.4</td>
<td>19.78</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>9.1</td>
<td>18.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>0.0</td>
<td>2.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>31.0</td>
<td>68.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child responding</td>
<td>M</td>
<td>29.5</td>
<td>33.9</td>
<td>1.07</td>
<td>.391</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>13.5</td>
<td>15.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>5.0</td>
<td>10.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>62.5</td>
<td>72.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child vocalizations</td>
<td>M</td>
<td>30.1</td>
<td>45.5</td>
<td>4.13</td>
<td>.058</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>17.8</td>
<td>30.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>1.6</td>
<td>11.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>80.8</td>
<td>127.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

reported that the father, after training, felt empowered in his paternal role and became an active school liaison. This was positively viewed by the mother and teacher and also proved beneficial for the child who now has both parents consistently involved in his education.

As previously discussed, children with autism have profound social and language deficits as well as generalization deficits that require special consideration when planning and implementing training procedures. Training methods used in this research build on earlier work (Calhoun et al., 1991; Elder, 1995; Mudford, Martin, Eikeseth, & Bibby, 2001; Mundy et al., 1994) provide support for social interaction theory and show great promise as strategies for promoting social relatedness and language development, two areas commonly deficit in children with autism. Results of this study also suggest the value of further characterization and evaluation of father–child interactions and father-directed in-home training. Also, more information is needed regarding the efficacy of additional specific theoretically linked training components to facilitate replication and development of comprehensive family-training procedures.

Accepted for publication August 10, 2004.
Funded by the National Institute of Nursing Research, R29 NR04297.
We thank Meredith Carlton for her sustained commitment and contributions related to project coordination and data management. We are also grateful to Gregory Cohen, Paula Modesitt, Tamara Diehl, and Lindsay Sherrill for their assistance in data collection and coding as well as Marnie Wiss and Sylvia Burns for their contributions related to preparing the final manuscript. Also critical to this work was research design and methodology consultation provided by Drs. Henry Pennypacker, Mark Lewis, and Brian Iwata. Last, and certainly not least, we acknowledge the autistic children and courageous families who continually inspire us in our work.

Corresponding author: Jennifer Harrison Elder, PhD, RN, FAAN, Box 100187, University of Florida College of Nursing, HPNP Building 101 Newell Road, Gainesville, FL 32611 (e-mail: elderjh@nursing.ufl.edu).
References


Midlife Women’s Adherence to Home-Based Walking During Maintenance

JoEllen Wilbur ▼ Annemarie Vassalo ▼ Peggy Chandler ▼ Judith McDevitt ▼ Arlene Michaels Miller

Background: Despite the many known benefits of physical activity, some women (27%) report no leisure-time physical activity in the prior month. Of those women who began an exercise program, the dropout rate was as high as 50% in the first 3–6 months. The challenge for researchers and clinicians is to identify those factors that influence not only adoption, but also maintenance, of physical activity.

Objective: The purpose of this study was (a) to describe midlife women’s maintenance of walking following the intervention phase of a 24-week, home-based walking program, and (b) to identify the effects of background characteristics, self-efficacy for overcoming barriers to exercise, and adherence to walking during the intervention phase on retention and adherence to walking.

Methods: There were Black and White women participants (N = 90) aged 40–65 years who completed a 24-week, home-based walking program. Self-efficacy for overcoming barriers to exercise, maximal aerobic fitness, and percentage of body fat were measured at baseline, 24 weeks, and 48 weeks. Adherence was measured with heart-rate monitors and an exercise log.

Results: Retention was 80% during maintenance. On average, the women who reported walking during maintenance adhered to 64% of the expected walks during that phase. Examination of the total number of walks and the number and sequence of weeks without a walk revealed dynamic patterns. The multiple regression model explained 40% of the variance in adherence during the maintenance phase.

Discussion: These results suggest that both self-efficacy for overcoming barriers and adherence during the intervention phase play a role in women's walking adherence. The findings reflect dynamic patterns of adopting and maintaining new behavior.

Key Words: adherence • exercise • maintenance • midlife • walking
Norcross (1992) have stressed, change is a dynamic process. Most people who take action to modify their behavior slip backward occasionally before they are successful. The stage of maintenance is not achieved until persons have engaged in a desirable behavior for 6 months. The current goal for a desirable level of physical activity is meeting the American College of Sports Medicine (ACSM) recommendation of vigorous physical activity 3 or more times per week for 20 min per session or engaging in regular (preferably daily) moderate physical activity at least 30 min per day (ACSM, 2000; Pate et al., 1995). For intervention studies, maintenance is typically defined as continued physical activity for at least 6 months after the completion of an intervention (Marcus et al., 2000). Even if participants successfully achieve physical activity at the desirable level identified by the ACSM by the end of an intervention, it is imperative to identify the dynamic patterns of adherence during maintenance in order to determine the effects of the intervention over time (USDHHS, 1996).

The research examining the maintenance of physical activity beyond the intervention phase is limited to mixed gender samples of Whites. In addition, it is difficult to compare across studies because of differing measures of adherence. On the basis of exercise-log data at the end of 1 year, midlife adults had higher adherence to both a home-based, high- and moderate-intensity exercise program than to a high-intensity, group-based program (78.7%, 75.1%, and 52.6%, respectively; King et al., 1995). Conversely, Cox, Burke, Gorely, Beilin, and Puddey (2003) found that midlife women who participated in a center-based exercise program had higher energy expenditure based on a 7-day recall 1 year later than women in a home-based exercise program. Only Bock, Marcus, Pinto, and Forsyth (2001) identified dynamic patterns of physical activity maintenance. In this study, over half of a sample of women participating in an intervention that compared two print-based intervention formats had a stable (but underactive) pattern of physical activity at 12 months (they did not achieve ACSM criteria at either 6 or 12 months). The validity of the patterns identified is compromised, however, by the use of a self-report recall measure of the number of minutes exercised in the prior week. According to Dishman (1994), self-report measures are the least accurate for persons who are sporadically active. Although exercise logs are also a self-report measure, they are filled out prospectively, providing multiple data points. They may be more useful in studying the dynamic nature of adherence.

Self-efficacy is pivotal in research on determinants of physical activity in previously inactive adults (McAuley & Blissmer, 2000). Specifically, exercise self-efficacy (the confidence in one's ability to be physically active under a number of specific circumstances) was a significant predictor of adoption and overall exercise levels in previously sedentary adult women enrolled in a 24-week, low-impact aerobic exercise class (McAuley & Jacobson, 1991). As reported earlier, adherence to a 24-week, home-based walking program was higher among those with a higher level of exercise self-efficacy at baseline (Wilbur, Miller, Chandler, & McDevitt, 2003). Although exercise self-efficacy at the end of an intervention has been examined less frequently, study results of healthy adults showed that it was predictive of individuals who maintain high levels of physical activity through follow-up (Bock et al., 2001); baseline self-efficacy was not predictive (Oman & King, 1998). Previous exercise behavior was a strong source of self-efficacy information often overlooked as a determinant of subsequent behavior (McAuley, Jerome, Elavsky, Marquez, & Ramsey, 2003).

In earlier reports, recent experience with exercise was the strongest predictor of subsequent participation in the maintenance phase of a group-based exercise in both middle-age and older men and women (McAuley, 1992; McAuley et al., 2003).

The study had two purposes:
1. to describe walking maintenance patterns for midlife Black and White women during 24 weeks after a walking intervention and
2. to identify how demographic characteristics (age, race, education, marital status, professional status), self-efficacy for overcoming barriers to exercise, and adherence to walking during the intervention phase affect retention and adherence to walking during a maintenance phase (weeks 24 through 48 after orientation).

Methods
Participants
Sedentary, healthy, employed, midlife Black and White women (45–65 years of age) residing in metropolitan Chicago were recruited to participate in a home-based walking intervention through oral and poster presentations at work sites and community churches, e-mail notices, flyers, newspaper notices, and an evening television news report. Sedentary was defined as not having exercised more than 20 min twice a week in the preceding 6 months. Healthy was defined as no more than one major coronary risk factor (in addition to age and sedentary lifestyle) and no major signs or symptoms of cardiopulmonary or metabolic disease. The exclusion criteria, which included use of psychotropic drugs, hormone replacement therapy, and a history of bilateral oophorectomy, were designed to control for influences other than walking on physical and psychological symptoms not reported here. Women with a body mass index (weight in kilograms/height in square meter) more than 40 were excluded because of the technical restrictions on the measurement of dual-photon x-ray absorptiometry (Wilbur, Miller, & Chandler, 2001). The women were initially screened for eligibility with a telephone questionnaire. Additional screening consisted of a health assessment with a nurse practitioner, a complete blood cell count and lipoprotein profile, calculation of percentage of body fat, and a maximal aerobic fitness test at the university Human Performance Laboratory.

Intervention and Procedure
The intervention was a home-based, moderately intense walking program (Wilbur et al., 2003). Each woman received an individually tailored exercise prescription that
Nursing Research  January/February 2005  Vol 54, No 1

Midlife Women's Adherence to Home-Based Walking During Maintenance

 included walking at a prescribed frequency, intensity, and duration. Participants walked 4 times per week within their target heart-rate range, which had been determined by a maximal aerobic fitness test. Participants gradually progressed to walking from 20 to 30 min in the target heart-rate range in addition to a 5-min warm-up and stretch and a 5-min cooldown and stretch. The number of walks was measured by both a self-report exercise log and a Polar Heart Rate Monitor during the 24-week intervention phase.

During the intervention phase, an assigned research-team member met with each woman every 2 weeks to deliver the intervention using sources of self-efficacy beliefs to enhance confidence for exercise, including mastery of exercise, verbal persuasion, and physiological arousal (Bandura, 1997). Mastery of exercise was addressed by reviewing exercise logs and heart-rate monitor data and providing emotional support and reinforcement in the form of feedback on the woman's progress. Verbal persuasion was provided by offering praise and encouragement, and by working with the woman to identify potential barriers and to develop plans for overcoming obstacles. Physiological arousal included transferring each woman's heart-rate monitor data to a notebook computer and showing her graphs of the intensity of her exercise. Also, a 2-week symptom questionnaire was administered so that each woman could monitor her physiological responses to exercise.

At the completion of the intervention phase, the women returned to the data-collection site to complete a questionnaire and physiological measures, including a maximal aerobic fitness test and calculation of percentage of body fat. They were given results of all physiological measures taken at baseline and at completion of the intervention. They were given also a new target heart-rate range on the basis of changes in their performance on the maximal aerobic fitness test, taught to take their pulse to monitor exercise intensity, and instructed to continue recording their walks in the exercise logs.

During the 24-week maintenance phase, a staff researcher called the participants monthly to remind them to mail or fax their exercise-log sheets. The women returned to the data-collection site at the end of the 24-week maintenance phase (48 weeks, or 12 months, after orientation) for administration of the final questionnaire and physiological measures, including a maximal aerobic fitness test and calculation of body fat. After completion of each of the three data collections, each woman received compensation for her time and travel.

Measures

Adherence to Walking Adherence during the intervention phase (Wilbur, Chandler, & Miller, 2001) was measured with a Polar Vantage XL heart-rate monitor and a self-report exercise log. Adherence to walking during the maintenance phase was measured by the self-report exercise log. The women recorded the date, time of completion of the warm-up/stretch and cooldown/stretch, and the number of minutes they walked for each session. The women were given credit during maintenance if they reported aerobic exercise other than walking. Adherence to the prescribed frequency was calculated for both the intervention phase and the maintenance phase as a percentage of the expected 96 walks (four walks per week for each 24-week period). The number of minutes walking for each session was not used because the majority of women completed the prescribed duration of 30 min per session. The correlation between the number of walks on the log and on the heart-rate monitor during the intervention phase was .96 (Wilbur, Chandler, et al., 2001).

Six dynamic, comprehensive, and mutually exclusive adherence patterns were identified over the intervention phase (Wilbur, Chandler, et al., 2001). These six mutually exclusive patterns were based on the total number of walks reported via both exercise logs and heart-rate monitors, and the number and sequence of weeks reported without any walks. For the present analysis, the original six patterns were modified and reduced to five patterns that described adherence in both the intervention and maintenance phases. A sixth pattern only for the maintenance phase was included.

1. Consistent adherence pattern completed 80% or more of the prescribed number of walks during the intervention or maintenance and had no lapses (lapse = 1 week with no walks).
2. Occasional lapse adherence pattern completed 80% or more of the expected number of walks and lapsed 1–4 weeks (during the intervention phase or maintenance) but had no relapse (relapse = 3 consecutive weeks with no walks).
3. Low-adherence pattern completed less than 80% of the prescribed walks, lapsed 0–4 weeks, but had no relapses during the intervention or maintenance phase.
4. Recyclers had a relapse but then returned to walking.
5. Relapsers had a relapse and did not walk again during the last 3 weeks of either the intervention or maintenance phase.
6. Drop did not complete the maintenance phase.

Two physiological measurements provided indirect measures of adherence including maximal aerobic fitness and percentage of body fat. Aerobic fitness was determined through a symptom-limited, incremental exercise test by using a treadmill and the modified Bruce-test protocol designed for use with participants of low fitness (American Heart Association Committee on Exercise, 1972; Pollock et al., 1982). The test began at Step I, 1.7 mph, 0% grade, and was followed by Step II, 1.7 mph, 5% grade. In each following stage, the grade increased 2%, and the speed increased either 0.8 or 0.9 mph until the participant was exhausted. The duration of each progressive stage was 3 min. Oxygen consumption and carbon dioxide production were measured from exhaled gases on a breath-by-breath basis using the 2900 Sensor Medic metabolic cart (ACSM, 2000). Heart rate was measured with a 12-lead electrocardiogram at rest and continuously during the test. Blood pressure was monitored by cuff sphygmomanometer at 3-min intervals. Percentage of body fat was calculated using the equation by Jackson, Pollock, and Ward (1980). Harpenden calipers (British Indicators Limited, England) were used to measure skinfold thickness in millimeters to the
nearest 0.1 mm at the triceps, suprailiac, abdomen, and thigh. Three measures were taken at each site on the right side of the body, averaged, and used to calculate percentage of body fat.

Retention in the Study  Retention differs from adherence in that it refers to the number of participants in the study at baseline, at 24 weeks (end of the intervention phase), and at 48 weeks (end of the maintenance phase).

Self-Efficacy  The 14-item McAuley Self-efficacy for Exercise Scale (McAuley & Jacobson, 1991) was used to measure exercise self-efficacy. The participant indicated confidence in her ability to continue walking in the face of barriers using a range from 10 (not confident) to 100 (completely confident). The items were summed and divided by the total number of items. Examples of barriers to exercise items included “weather very bad” or “felt pain or discomfort when exercising.” Three additional items identified from the literature as exercise barriers specifically for women were added to the original measure. The additional items were “family relationships and responsibilities,” “lack of encouragement from family,” and “personal safety.” The resulting scale was composed of 17 barriers to exercise. The Cronbach alpha for the scale was .92.

Analyses  First, differences between the women who completed the maintenance phase and the women who did not were examined using chi-square (for categorical variables) test and two-sample t (for numeric variables) test. Differences between the two groups in demographics (i.e., age, race, education, marital status, professional status), adherence to walking, and baseline to 24-week changes in self-efficacy (change in exercise self-efficacy score) and physiological measures (percent changes in maximal aerobic fitness and percent body fat) were examined. All tests used an alpha level of .05.

Second, the effects of adherence to walking during the intervention phase, exercise self-efficacy, physiological measures, and background characteristics on adherence to walking during the maintenance phase were evaluated using multiple regression analyses. Finally, differences among patterns of adherence during the intervention phase and patterns of adherence during the maintenance phase were examined.

Results  Of the 173 women who met the screening criteria, 102 women were randomly assigned to a 24-week walking intervention phase, to be followed by an additional 24-week maintenance phase. Of the 102 women, 90 returned for their 24-week (6-month) measures for a retention rate of 88% in the intervention phase; their data are presented in this study. The women ranged in age from 45 to 63 years (M = 51.3, SD = 9.1). The sample was 30% Black and 70% White; 53% of participants were married, and the majority of women (66%) had 1–3 children. A majority of participants reported having a college degree (68%), whereas 63% held professional positions.

Of the 90 women who completed the intervention phase, 72 also completed the maintenance phase for a retention rate of 80%. The reasons women dropped out during the intervention phase (n = 18) included medical problems (8), no time (5), family problems (3), relocation (1), and no reason provided (1). The final contact with the women dropping out was made by a staff researcher who contacted 7 of the women in the first 2 months and the remaining 11 women in the last month of the maintenance phase. During the final contact, all of the women reported that they had not exercised following the completion of the intervention phase.

Examination of the exercise-log data for the women who did complete the maintenance phase indicated that they adhered to 64.6% (SD = 0.286, range = 6%–100%) of the expected frequency of 96 walks (4 times a week for 24 weeks) during the maintenance phase. The average duration per walk was 39.30 (SD = 9.24) min, which was longer than the prescribed duration of 30 min per session. The average number of minutes women spent walking ranged from 20 to 56.9.

A chi-square test showed that the women who did not complete the maintenance phase differed from the women who completed maintenance by race, but not by marital status, professional status, or education (Table 1). Retention was higher for White women than for Black women, 86% and 67%, respectively.

A two-sample t test showed significant differences in percentage of adherence during the intervention, exercise self-efficacy at the end of the intervention, and change in exercise self-efficacy and body fat from baseline to the end of the intervention between the women who completed maintenance and those who did not (Table 2). The women who completed the maintenance phase had greater adherence (81% vs. 47%) during the intervention and higher scores on exercise self-efficacy (71% vs. 54% confidence) to overcome barriers at the end of the intervention than the women who did not complete maintenance phase. In addition, the change in exercise self-efficacy from baseline to 24 weeks was greater than the change in those women who did not complete the maintenance phase. The women who completed maintenance had a greater loss in percentage of body fat from baseline to the end of the intervention phase than the women who did not participate in walking. Women who completed maintenance had a greater increase in maximal aerobic fitness from baseline to the end of the intervention phase than the women who did not participate in walking (p = .056).

Forty percent of variance was explained when percent adherence to walking during the maintenance phase as the outcome variable was regressed on (a) percent adherence during the intervention phase, (b) self-efficacy for overcoming barriers at the end of the intervention (24 weeks), (c) change scores for self-efficacy and percent change for maximal aerobic fitness and percent body fat from baseline to the end of the intervention, and (d) race as independent variables (Table 3). Both adherence to walking during the intervention phase and self-efficacy at the end of the intervention were significant predictors of adherence to walking during the maintenance phase. Women who had greater adherence to walking during the intervention phase and
higher exercise self-efficacy scores at the end of the intervention had greater adherence during maintenance. The women were categorized into adherence patterns on the basis of the number of walks they completed and relapses they had during the maintenance phase (Table 4). There were 18 women (20%) in the consistent adherence pattern, 10 women (11%) in the occasional lapse adherence pattern, and 8 women (8%) in the low-adherence

<table>
<thead>
<tr>
<th>TABLE 1. Descriptive Statistics for Background Characteristics by Completion of Maintenance (N = 90)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Background Characteristics</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Race</td>
</tr>
<tr>
<td>Black</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Marital status</td>
</tr>
<tr>
<td>Married</td>
</tr>
<tr>
<td>Not married</td>
</tr>
<tr>
<td>Professional status</td>
</tr>
<tr>
<td>Professional</td>
</tr>
<tr>
<td>Nonprofessional</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>College graduate</td>
</tr>
<tr>
<td>Not college graduate</td>
</tr>
<tr>
<td>Age, ( M (SD) )</td>
</tr>
</tbody>
</table>

Note. ns = Not significant.

<table>
<thead>
<tr>
<th>TABLE 2. Summary Statistics of Adherence to Walking, Self-Efficacy, Maximal Oxygen Uptake, and Percent Body Fat by Completion of Maintenance (N = 90)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Adherence (%)</td>
</tr>
<tr>
<td>During intervention</td>
</tr>
<tr>
<td>Self-efficacy (%)</td>
</tr>
<tr>
<td>Baseline</td>
</tr>
<tr>
<td>End of intervention</td>
</tr>
<tr>
<td>Change</td>
</tr>
<tr>
<td>Maximal aerobic fitness (%)</td>
</tr>
<tr>
<td>Baseline</td>
</tr>
<tr>
<td>End of intervention</td>
</tr>
<tr>
<td>Percentage Change</td>
</tr>
<tr>
<td>Body fat (%)</td>
</tr>
<tr>
<td>Baseline</td>
</tr>
<tr>
<td>End of intervention</td>
</tr>
<tr>
<td>Change</td>
</tr>
</tbody>
</table>

Note. ns = Not significant.
Twelve women (13%) were classified as recyclers and 24 (27%) as relapers. The drop pattern included the 18 (20%) nonexercising women who did not complete maintenance. Examination of adherence patterns at the end of the intervention phase versus those at the end of the maintenance phase showed that 24% had not changed their walking patterns, 12% of the women had moved to a pattern that suggested an increase in their walking, and over half (63%) had moved to a pattern that involved a decrease in their walking. All of the women in the top two adherence patterns at the end of the intervention phase reported walking during the maintenance phase.

**Discussion**

Of the 102 women who began the walking intervention, a majority (71%) completed the maintenance phase. This is higher than the 65% retention of men and women participating in an intervention comparing mailed tailored counseling messages versus mailed self-help booklets promoting physical activity (Bock et al., 2001), and the 62% of women participating in either a low-intensity or a high-intensity, home-based exercise program at the end of the maintenance phase (12 months; Cox et al., 2003). Retention at 12 months for this home-based walking program was substantially higher than the 50% retention previously suggested for group-based programs (Dishman, 1990). Retention was lower, however, than the overall 82% retention rate reported for women participating in an intervention comparing high-intensity, group-based; high-intensity, home-based; and lower-intensity, home-based exercise programs conducted in California (King et al., 1991). Although none of the participants in this study mentioned climate as a contributing factor to dropping from the walking program, climate was anecdotally reported to be problematic throughout the winter months. King and colleagues included predominately White men and women, which may have also contributed to higher participation in that study. All of the women in this study were employed as opposed to 64.7% of the women in the earlier study (King et al., 1995), which suggests an additional time burden in this sample.

The retention rate was higher when the time period from the end of the intervention phase to the end of maintenance phase at 1 year was examined. Eighty percent of the women who participated in the active walking intervention continued to walk and complete the maintenance phase.
phase, similar to 72% reported by Cox et al. (2003) and 80% reported by Bock et al. (2001). As in prior studies (Cox et al., 2003), the most frequent reason for withdrawing from the walking program was medical problems. These problems varied greatly, with only one woman reporting a musculoskeletal problem related to walking. The low incidence of musculoskeletal problems reflects the advantages of a moderate-intensity prescription over a high-intensity prescription. The next most frequent reason for withdrawing was a lack of time (e.g., work schedules, family events, and crises). This is consistent with cross-sectional studies that report the lack of time as a main reason for lack of participation in physical activity (Eyler et al., 2002).

Prior studies document that some participants may be retained but they may walk at suboptimum levels. In this study, mean percentage of frequency of walks during maintenance was less (64%) than that during the intervention phase (80%), but it compares favorably to the findings of Jakicic et al. (1999), who reported a 47.2% adherence to walking 150 min. King et al. (1999), however, using exercise logs and the same definition of adherence as used in this study, describes a 75.1% adherence to a home-based, moderate-intensity exercise program after 12 months.

The mean duration of walks (39.3 min) exceeded the expectations of 30 min (Wilbur, Chandler, et al., 2001). It appears that the most challenging aspect of exercise adherence, both during the intervention and in maintenance, is getting started. Once women actually get out and walk, they meet or exceed the prescribed duration.

One of the purposes of this study was to examine the influence of background characteristics and exercise self-efficacy on retention and adherence to walking during maintenance in midlife Black and White women. Like Bock et al. (2001), we found no significant differences between participants and nonparticipants in the maintenance phase on baseline education, age, or body fat. However, we identified that a higher proportion of Black than White women did not participate in or complete the maintenance phase. This disparity warrants further investigation. As reported earlier, a generic videotape and materials that did not include Blacks were used to orient the women to the walking program (Wilbur et al., 2003). The results highlight the importance of developing intervention materials that are culturally sensitive to the targeted population.

Both higher self-efficacy for exercise and improved self-efficacy at the end of the intervention phase as well as greater adherence to frequency of walking during the intervention phase distinguished women who completed the maintenance phase from those who did not and influenced adherence to frequency of walking during maintenance. These findings are similar to those of other studies showing that individuals who continued to maintain high levels of physical activity through the follow-up period had higher self-efficacy and were less likely to perceive barriers to physical activity at the end of the intervention (Bock et al., 2001; King et al., 1991; McAuley & Jacobson, 1991; Oman & King, 1998). Also, similar to prior research (Oman & King, 1998), this study showed that baseline exercise self-efficacy did not predict exercise adherence during the maintenance phase. Improved self-efficacy from baseline to the end of the intervention phase influenced adherence during the maintenance phase. This suggests that self-efficacy is dynamic. The influence of change in self-efficacy on retention reinforces the importance of targeting interventions toward improving self-efficacy for exercise.

Frequency of walking at the end of the intervention was the strongest predictor of subsequent walking during the maintenance phase. It appears that recent exercise experience or mastery could be a source of self-efficacy information, which influences adherence during maintenance.

The women who completed the maintenance phase had a loss in body fat during the intervention phase whereas the women who did not complete the maintenance phase on average gained body fat during the intervention. It may be important to include a measure of percentage of body fat rather than body weight alone, because body weight may not change with moderate exercise. For example, Dunn and colleagues (1999) found no change in body weight in men and women exercising at 50%-85% intensity 3-5 days a week for 6 months, but found that body fat decreased by 1.85% on average.

Examination of the exercise patterns revealed that 31% of the women followed either a continuous or an occasional lapse pattern through the maintenance phase of the study and a much larger percentage (63%) regressed to a less active pattern. The majority of women who did not walk reported health-related problems. These findings further reinforce the need for additional time and verbal persuasion targeted for women who have been forced to recycle to a less active lifestyle.

The findings of this study indicate that physical activity motivation, or self-efficacy, may be increased through verbal persuasion by staff researchers, by developing awareness of physiological responses to activity, and through the performance accomplishments of increased physical activity. Evidence of this is shown in the women who completed the intervention phase with higher self-efficacy toward overcoming specific barriers to exercise. Through their 24-week experience with the intervention, they had developed effective personal strategies supporting regular physical activity. They continued to use these strategies during maintenance and thus were able to adhere, resulting in ongoing validation of their ability to devise effective strategies. Successful deployment of these strategies resulted in continuing experiences of physiological arousal and ongoing accrual of performance accomplishments. ▼

Accepted for publication August 26, 2004.

This study was supported by a grant from the National Institute of Nursing Research, National Institutes of Health (RO1 NRO4234). The authors thank Edward Wang, PhD, Research Assistant Professor, College of Nursing, University of Illinois at Chicago, for his assistance in the statistical analyses for this article.

Corresponding author: JoEllen Wilbur, PhD, RN, FAAN, Department of Public Health, Mental Health and Administrative Nursing, College of Nursing, University of Illinois at Chicago, Chicago, Illinois 60612 (e-mail: jwilbur@ucc.edu).
References
Predictors of Professional Nursing Practice Behaviors in Hospital Settings

Milisa Manojlovich

Background: Many hospital nurses perform isolated, routine tasks, rather than use their professional training, because they are subject to control by organizational and medical divisions of labor. The environment may interfere with a nurse’s ability to practice autonomously and according to professional standards.

Objectives: The purpose of the study was to explore how certain factors in the environment and personal characteristics interact to affect hospital nursing practice behaviors.

Methods: The study used a nonexperimental, comparative design. Surveys were sent to a random sample of 500 nurses throughout the state of Michigan. Three instruments, measuring structural empowerment, self-efficacy for nursing practice, and professional practice behaviors, were included. Path analysis was used for statistical analysis.

Results: Three hundred sixty-four nurses responded (73%), of whom 251 provided usable protocols for the final analysis. Environmental factors (structural empowerment) contributed both directly to professional practice behaviors as well as indirectly through self-efficacy. Self-efficacy mainly exerted its effect as a mediator in the relationship between environmental factors and practice behaviors. Support for the proposed theoretical model was mixed, although the proposed model fit the data well ($\chi^2 = 11.02 \ (d.f. = 5, \ N = 251), \ p < .05, \ CFI = .999, \ NNFI = .991, \ RMSEA = .069$). An alternative model emerged from the data analysis.

Discussion: Nurses may practice more professionally when the environment provides opportunities and power through resources, support, and information. Self-efficacy may contribute to professional practice behaviors, especially in an environment that has the requisite factors that provide empowerment.

Key Words: professional nursing practice behaviors • self-efficacy • structural empowerment

Many hospital nurses perform isolated, routine tasks instead of higher level skills that require independent judgment and thinking (Shorr, 2000). Shorr commented on the loss of a professional focus in nursing, stating, “the behavior of many nurses can only be described as ‘mechanical’” (p. 90). Understanding the nature of nursing practice behaviors is important for many reasons, but two reasons in particular are relevant. First, the current nursing shortage suggests that not only have nurses been highly dissatisfied with their positions (Cumbey & Alexander, 1998) but also the profession of nursing has not been as attractive a career choice as it once was (Buerhaus, Staiger, & Auerbach, 2000). The task-centered focus of much nursing work may be in part to blame, and one answer to the nursing shortage may lie in providing opportunities for a more variable and autonomous job. Second, nursing practice, as exemplified by a focus on task-centered behaviors, may also adversely affect patient outcomes. Research has shown that patient outcomes improved when the hospital organization supported professional practice characteristics, rather than task-centered behaviors (Aiken, Clarke, & Sloane, 2000).

Three essential attributes of professional nursing practice were identified by Scott, Sochalski, and Aiken (1999) in their review of the original magnet hospital research as well as other descriptive studies, which used the magnet hospital framework. The attributes included (a) the ability of the nurse to establish and maintain therapeutic relationships with patients (usually through a primary nursing delivery model); (b) nursing autonomy and control over the practice environment; and (c) collaborative relationships with physicians at the nursing-unit level (Scott et al., p. 10). Aiken and colleagues later conceptualized autonomy, control over the practice environment, and nurse-physician collaboration as hospital characteristics (Aiken, Havens, & Sloane, 2000).

Milisa Manojlovich, PhD, RN, CCRN, is Assistant Professor, School of Nursing, University of Michigan, Ann Arbor.
It was likely that nurses frequently were unable to use their professional training, training that focused on autonomous practice and independent decision making, because they were controlled by both organizational and medical divisions of labor (Alexander, 1982). Without the authority to make decisions, nurses may not be able to work effectively, “in a manner consistent with professional standards” (Alexander, p. 21). Providing decision support to nurses at the point of care has more recently been identified as promoting patient safety (Page, 2004).

Organizational systems that encourage nurses to use their professional skills may provide a rewarding career choice for current students (Bednash, 2000). Theoretical (Kanter, 1993) and empirical studies suggest that when the organization provides opportunity and access to authority, employees are more satisfied and more effective on the job (Laschinger & Havens, 1997; Laschinger, Shamian, & Thomson, 2001). Providing bedside nurses with control over both the content and the context of their practice may increase their involvement in decision making (Laschinger, Sabiston, & Kutscher, 1997), which has been deemed essential to professional practice (Alexander, 1982). However, a continued lack of control over the content and context of their work would suggest that power remains an elusive attribute to many nurses. The complex issue of task-centered (instead of profession-centered) behavior should be investigated. Therefore, the purpose of this study was to explore how certain factors in the environment and personal characteristics interact to affect nurse practice behaviors.

Review of the Literature

Structural empowerment is a concept that refers to four social structure factors in the environment that, when available, promote employee effectiveness and satisfaction. These factors are opportunity, information, resources, and support (Kanter, 1993). Structural empowerment has been studied in the nursing literature and linked to many nurse outcomes, some of which may be associated with professional practice behaviors. Sabiston and Laschinger (1995) demonstrated that a significant, positive relationship existed between job-related empowerment and nurse autonomy. In a secondary analysis of two earlier studies, it was demonstrated that nurses perceived significantly greater control over both the content and the context of their practice when they had access to work empowerment structures (Laschinger et al., 1997). Earlier autonomy and control over the practice environment had been identified already as hallmarks of professional nursing practice (Schutzenhofer & Musser, 1994).

Very little published research has examined the concept of self-efficacy for nursing practice, although nursing has studied self-efficacy as it relates to nurse managers, nursing students, and patient care (Haas, 2000). Self-efficacy has been defined as the belief in one’s abilities to mobilize the motivation, cognitive resources, and courses of action needed to exercise control over one’s work (Bandura, 1997). Caring has been identified as being pivotal to the nursing role as well as being the essence of nursing (Wolf, Giardino, Osborne, & Ambrose, 1994). Caring self-efficacy has been defined as “nurses’ beliefs in their abilities to express caring orientations, attitudes, and behaviors and to establish caring relationships with clients or patients” (Coates, 1997, p. 54). Therefore, caring self-efficacy may be appropriate to use as an indicator of self-efficacy for nursing practice.

Self-efficacy has been linked to empowerment as a crucial stage of the empowerment process (Kramer & Schmalenberg, 1993), suggesting that a relationship may exist between empowerment and self-efficacy. Access to both information and resources—two components of structural empowerment—has been found to enhance self-efficacy as well (Spreitzer, 1996). Laschinger and Shamian (1994) found a positive relationship between managerial self-efficacy and job-related empowerment, suggesting a possible link between organizational structures and personal levels of efficacy necessary to carry out managerial roles.

Although the relationship between professional nursing practice behaviors and self-efficacy has not previously been investigated, a few studies have suggested that there may be a relationship between these two concepts. Self-efficacy beliefs have been demonstrated to aid in efficient analytic thinking in complex decision-making situations (Bandura, 1989). In addition, perceived self-efficacy has been found to contribute to higher levels of productivity and improved performance (Gist & Mitchell, 1992).

Research has shown also that personal self-efficacy determines the levels of motivation by determining the amount of effort to exert on a task or how long to persist in the face of obstacles (Bandura, 1989). Considering the complex nature of modern health care environments, nurses may benefit from increased levels of self-efficacy. To identify the appropriate course of action and function effectively, professionals must have an understanding of and control over the activities associated with their job (Alexander, 1982) as well as a belief in their ability (Bandura, 1997). Wide variability in reaction to the same environment occurs because individuals are influenced by their perceptions of the environment, and not objectivity (Spreitzer, 1996). Thus, a nurse’s perception of the work environment and a sense of self-efficacy are important to understand variation in practice behaviors.

Research Aims and Study Questions

The following specific aims and related research questions were proposed for study:

Aim 1: To examine the effects of social structural factors in the work environment, known collectively as structural empowerment, on professional nursing practice behaviors. What is the relationship between structural empowerment and professional nursing practice behaviors?

Aim 2: To examine the effects of a personality variable, self-efficacy, on both environmental factors (structural empowerment) and professional nursing practice behaviors. What is the relationship between self-efficacy and structural empowerment? What is the relationship between self-efficacy and professional nursing practice behaviors?

Theoretical Framework

The theoretical framework developed for this study blended two well-established theories to explain variation
in professional nursing practice. The theory of structural empowerment (Kanter, 1993) recognized the environment for its ability to influence behavior whereas Social Cognitive Theory (SCT) maintained that interpretation of the environment was dependent, in part, upon self-efficacy (Bandura, 1997). These theories provided two distinct perspectives, because they have shown variation in human work behavior arising from different causes.

Both the work environment and the manner in which it is interpreted affect nursing practice behaviors. Some environments are empowering because they allow workers flexibility in getting the job done. Other work environments may not be empowering, yet there are individuals who manage to be effective on the job.

It may be that these individuals believe in their ability to mobilize the necessary psychological resources to complete the job. Many individuals believe themselves to be effective in multiple spheres of influence (e.g., parents, wood carvers, cooks, car drivers). Yet, in the workplace, these same people no longer believe they can be effective. On one hand, the hospital environment, or any work environment, without empowering structures, is likely to diminish perceptions of self-efficacy over time. On the other hand, a certain amount of self-efficacy may be necessary to be able to use empowering structures that are present. Therefore, the relationship between empowerment and self-efficacy may be reciprocal. The interaction between environmental structures, as expressed by structural empowerment, and the psychological attribute of self-efficacy may explain variation in nursing practice, from a focus on tasks at one end of the spectrum to practice behaviors more consistent with professional standards at the other. On the basis of the theoretical framework described earlier, a conceptual model was tested in this study (Figure 1).

Methods

Design

This descriptive study used a comparative survey design. Approval was obtained from the Institutional Review Board (IRB) before the study was initiated.

Variables and Instruments

Structural empowerment: Structural empowerment was measured by three empowerment scales: the Conditions for Work Effectiveness Questionnaire-II (CWEQ-II), the Job Activities Scale II (JAS-II), and the Organizational Relationships Scale II (ORS-II). All scales measured various aspects of Kanter’s concept of empowerment (Kanter, 1993). The CWEQ-II is a 12-item, four-subscale measure. The four subscales are Opportunity, Information, Support, and Resources. The JAS-II is a three-item measure of Kanter’s concept of formal power whereas the ORS-II measures Kanter’s concept of informal power. All three scales, 19 items in total, have demonstrated high internal consistency in multiple studies, ranging from .78 to .93 (Laschinger, Almost, & Tuer-Hodes, 2003). Content and construct validity of the CWEQ-II have been established (Laschinger, Finegan, Shamian, & Wilk, 2001) and the factor structure of both the JAS-II and ORS-II scales have been validated. A total empowerment score can be created by summing all six subscales (range = 6–30). Only the total score was used in this analysis.

Self-efficacy: The Caring Efficacy Scale (CES) was used to measure caring self-efficacy, and is a 30-item self-report tool, arranged in a 6-point Likert-type format (Coates, 1997). Items are balanced between positive and negative content. Scores are summed and averaged, with higher numbers associated with higher efficacy beliefs. The Cronbach’s α reliability coefficients have ranged from .85 to .92 (Meretoja & Leino-Kilpi, 2001). The scale has been tested for content (via expert groups) and concurrent validity (Meretoja & Leino-Kilpi), although it has yet to be subjected to factor analysis (Coates). Sample scale items of the CES include, “I can usually create some way to relate to most any client/patient,” and “Clients/patients can tell me most anything and I won’t be shocked.”

Professional nursing practice: Professional nursing practice was measured by the Nursing Activity Scale (NAS). With this 30-item self-report 4-point Likert-type scale, scores can range from 60 to 240 and are based on the sum of weighted item scores. Continuing work with the NAS has yielded α reliability coefficients of .81 to .92

Sample and Procedures

The sample consisted of 500 nurses randomly selected from a list of 1,509 names provided by the Michigan Nurses Association (MNA). A list of medical—surgical nurses was specified from the MNA to achieve a more homogeneous sample. Of the 365 individuals who responded, 308 provided usable surveys and demographic information and signed informed consent forms. Ten of the 308 surveys were completed by nurses who did not work in hospitals. These 10 cases were dropped from the sample, because the study purpose was to investigate the effect of environmental and personal influences on hospital nurses only. To achieve further sample homogeneity, managers, supervisors, and nursing faculty were not included in the final sample; thus, an additional 32 cases were dropped.

Dillman’s (2000) method of implementing surveys was adapted for this study. The survey responses were self-reported in paper-and-pencil instruments. Three contacts were made to obtain responses.

FIGURE 1. Initial theoretical model.
(Schutzenhofer & Musser, 1994) as well as test-retest reliability \((r = .79)\). Content validity was established during the initial item generation process and after revision by consulting with nursing experts. Construct validity was established through factor analysis in which three strong factors emerged: individual development of the nurse; development of the professional role; and development of the role of patient advocate (K. Kelly, personal communication, October 13, 2002). A summary of all instruments and \(\alpha\) coefficients for reliability for this study is provided in Table 1.

Control variables: Because educational level (Johnson, 1988) and years of work experience (McCloskey & McCain, 1988) have been associated with more professional nursing practice, these variables were added to the model as control variables. No study has examined specialty certification’s effect on practice behaviors, although a recent study that emerged: individual development of the nurse; development of the role of patient advocate (K. Kelly, personal communication, October 13, 2002). A summary of all instruments and \(\alpha\) coefficients for reliability for this study is provided in Table 1.

Instrumental variable: An instrumental variable is related to one variable, but not the other, in a reciprocal relationship (Heise, 1975). Global empowerment was chosen as the instrumental variable because of its strong relationship to structural empowerment in previous research and because, theoretically, it was expected to be less related to self-efficacy. Global empowerment was created from two additional items of the CWEQ, separate from and not included in the total scale. Because its only purpose was to be able to correctly identify the nonrecursive model for path analysis, global empowerment was not otherwise included in any analysis.

Data Analysis

Data analysis was performed using the Statistical Package for the Social Sciences (SPSS, 2001), version 11.0, and Analysis of Moment Structures (AMOS) statistical software programs (Information Technology Services [ITS], 2001). Descriptive analyses of the study sample and variables were conducted. Inferential statistics included correlations, reliability assessments of study instruments, Sobel’s tests, and path analysis. Path analysis, a causal modeling technique, was used to test the theoretical model presented in Figure 1. To determine whether self-efficacy acted as a mediator in the relationship between empowerment and professional practice behaviors, both path analysis and Sobel’s tests were performed. Baron and Kenny (1986) recommend use of Sobel’s test to determine significant mediating effects. The level of significance chosen for this study was .05.

Results

The nurses in the sample ranged in age from 21 to 75 years \((M = 45.42)\), had an average of 20 years’ experience in nursing, and had spent more than 10 years in both their positions \((M = 10.27)\) and institutions \((M = 15.02)\). The sample consisted of individuals who were mainly female \((96.2\%)\), and White \((92.1\%)\). Most participants were educationally prepared at either the associate \((36.8\%)\) or baccalaureate \((37.6\%)\) level, with the remainder being diploma \((14.3\%)\), master’s \((10.5\%)\), or doctorally prepared \((0.8\%)\) nurses. Most nurses worked full-time \((59.8\%)\).

A correlation matrix (Table 2) was generated to begin an understanding of the relationships among study variables. The NAS scale used to measure professional practice behaviors was moderately related to both the empowerment \((r = .32, p < .01)\) and self-efficacy \((r = .45, p < .01)\) scales, suggesting that both structural empowerment and self-efficacy were positively related to professional practice behaviors. Of the control variables, education \((r = .26, p < .01)\) and specialty certification \((r = .17, p < .01)\) were also significantly related to professional practice behaviors.

All cases with any missing data were deleted, with the result that 251 cases were used to test the model. The theoretical model was nonrecursive, because of the possible reciprocal relationship between structural empowerment and self-efficacy. Therefore, to identify the model (Heise, 1975), an instrumental variable (global empowerment) was used. As predicted, global empowerment was highly correlated to the total structural empowerment score \((r = .77, p < .01)\), but not significantly related to self-efficacy \((r = .11)\).

Testing the theoretical model involved several stages. In the first stage of model testing, a model was generated that consisted of the main variables of interest (structural

<table>
<thead>
<tr>
<th>TABLE 2. Pearson Correlation Coefficients for Study Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>NAS</td>
</tr>
<tr>
<td>CES</td>
</tr>
<tr>
<td>CWEQ</td>
</tr>
<tr>
<td>Specialty certification</td>
</tr>
<tr>
<td>Years' experience</td>
</tr>
</tbody>
</table>

Note. NAS (Nurse Activity Scale) measures professional practice behaviors. CES (Caring Efficacy Scale) measures self-efficacy. CWEQ (Conditions for Work Effectiveness Questionnaire) measures structural empowerment.

\(*p < .01\) level (two-tailed test).
empowerment, self-efficacy, and professional practice behaviors) as well as control variables (type of nursing education, specialty certification, and years of work experience).

To determine model adequacy, the model was evaluated for its goodness of fit to the data (Kline, 1998). The Bentler Comparative Fit Index (CFI) indicated the relative proportion in improvement of the overall fit of the proposed model to a null model (Kline). Values for the CFI range from 0 (poor fit) to 1 (perfect fit); higher values are desirable. The Bentler–Bonett Non-Normed Fit Index (NNFI), also known as the Tucker–Lewis Index (TLI), is similar to the CFI, but penalizes for model complexity. Values close to 1 indicate good fit. Finally, Steiger’s Root-Mean-Square Error of Approximation (RMSEA) is an index that penalizes for lack of model parsimony; lower values are desirable. In addition, values < .06 are considered indicative of good model fit (Kline). This model fit the data well (Figure 2), as demonstrated by a χ² of 11.02 (df = 5, N = 251), p < .05, CFI = .999, NNFI = .991, RMSEA = .069).

In model testing, path significance is based on values of the critical ratio (cr), which is the ratio of the unstandardized parameter estimate to the standard error of that estimate. Critical ratios >1.64 were considered to be significant, using a one-tailed test because a direction was proposed for each relationship. Both structural empowerment (β = .20, cr = 3.76) and self-efficacy (β = .40, cr = 7.48) were significant predictors of professional practice behaviors. In fact, self-efficacy was a stronger predictor of practice behaviors than structural empowerment. The path from structural empowerment to self-efficacy was significant as well (β = .15, cr = 1.74). However, the path from self-efficacy to structural empowerment was not significant (β = .02, cr = .32).

To determine whether self-efficacy mediates the relationship between structural empowerment and professional practice behaviors, additional analyses were conducted. A new model was generated wherein the path from self-efficacy to structural empowerment was removed, as was the instrumental variable and its effect on structural empowerment.

An examination of critical ratios in all three paths demonstrated that the direct path from structural empowerment to professional practice behaviors was significant (cr = 3.73), as were both indirect paths (the path from structural empowerment to self-efficacy [cr = 2.63] and the path from self-efficacy to professional practice behaviors [cr = 7.48]). All component paths have to be significant in order for indirect effects and direct effects to be considered significant (Kline, 1998). However, the product of indirect paths was not greater than the direct path. These results indicate that the direct impact of structural empowerment on practice behaviors was greater than its indirect impact. To demonstrate complete mediation, the previously significant direct path would have to become insignificant (Baron & Kenny, 1986).

A series of equality-constrained models were generated as an additional test to determine if self-efficacy mediated the relationship between structural empowerment and professional practice behaviors. The basic mediator model without any path constraints was just identified, and therefore, had no degrees of freedom and a χ² equal to zero. In the model created to test mediation, the paths from structural empowerment to self-efficacy and from self-efficacy to professional practice behaviors were both constrained to zero at the same time. When this model was compared to the model without any path constraints imposed, the χ² difference was significant (χ² = 57.45, df = 2), indicating that there were indirect paths through self-efficacy and that self-efficacy did mediate the relationship between structural empowerment and professional practice behaviors.

The Sobel test, performed as an additional validation of the mediation effect (Preacher & Leonardelli, 2001), was significant (p = .013). The results of all models tested to this point therefore failed to support the theoretical model. However, an alternative theoretical model (Figure 3) emerged from the data analysis.

**Discussion**

The findings indicate that structural empowerment contributed to professional practice behaviors directly as well as indirectly through self-efficacy. The results failed to support the notion of a reciprocal relationship between structural empowerment and self-efficacy. A stronger contribution from self-efficacy to professional behaviors than from structural empowerment to professional practice behaviors are also suggested. The contribution of self-efficacy to professional practice behaviors indicated a moderate effect, whereas structural...
empowerment’s contribution to practice behaviors indicated a small to moderate effect. Path coefficient values < .10 may indicate a small effect, values around .30 a medium effect, and values > .50 a large effect (Kline, 1998).

The model predicted 30% of the variance in professional practice behaviors and provided mixed support for the theoretical model. Both structural empowerment and self-efficacy directly and significantly contributed to practice behaviors, and, further, structural empowerment contributed directly and significantly to self-efficacy. The findings are aligned with path analysis showing that structural empowerment is not only related to, but may cause, professional behaviors, adds more information about the outcomes of structural empowerment. Laschinger and Havens (1996) were able to demonstrate relationships between structural empowerment and work effectiveness as well as between structural empowerment and perceived control over nursing practice. Work effectiveness may be viewed as a positive contributor to professional behaviors, whereas control over nursing practice is one of the characteristics of professional practice behaviors used in this study.

The significant relationship extending from self-efficacy to professional behaviors is important because it provides information about the determinants of professional practice behaviors and about the sources of those determinants. The current research has concluded that environmental and personal factors influenced the development of professional practice behaviors, and is consistent with SCT (Bandura, 1997).

The findings show that self-efficacy partially mediated the relationship between structural empowerment and professional behaviors, and are consistent with Kanter’s theory of structural empowerment (Kanter, 1993). The indirect, mediating effect of self-efficacy on the relationship between structural empowerment and professional practice behaviors may be responsible for the relative smaller direct effect of structural empowerment on practice behaviors. Baron and Kenny (1986) assert that as a mediator effect becomes greater, the direct relationship between independent and outcome variables becomes smaller, with full mediation occurring when the relationship between independent and outcomes variables is zero.

Although no previously published study has examined the relationship between structural empowerment and self-efficacy for nursing practice, structural empowerment may precede and contribute to self-efficacy, as was found in the present study. Hospitals that provide opportunity, information, resources, and support may engender staff nurses’ beliefs that they can exercise control over their work lives. Theoretical evidence of this relationship has been established by Laschinger (1996).

Several limitations were identified in this study. Concepts that are processes rather than static, isolated occurrences were measured with a single self-assessment. Longitudinal study would better determine if these relationships exist over time. Several measures were taken to reduce social desirability response bias that can occur when participants rate their own behaviors. For example, participants were assured that their identity would be protected. Consistency artifact was minimized by varying the order of instruments placed in each packet (Podsakoff & Organ, 1986). Using a list from the MNA is another limitation, because it could be argued that nurses who choose to belong to a professional organization are different from those who do not. However, many hospitals in Michigan require MNA membership because the MNA is the bargaining unit, which offsets this limitation somewhat.

On one hand, organizational changes to improve the hospital work environment for nursing practice may be simplified by using structural empowerment as a blueprint for improvement. On the other hand, it may be that the impetus for a satisfactory work environment has to be initiated by staff nurses, rather than by hospital administrators. Such an approach has been recommended by the American Academy of Colleges of Nursing (Miller et al., 2002), highlighting the importance of the practice environment to professional nursing practice. Nurses may be able to actively seek work environments that foster more professional practice behaviors by asking about structural empowerment factors during the job interview process. Alternatively, nurses who are established in a particular hospital setting may conduct an environmental assessment of their work environment, appraising structural environment factors that may already be present, but previously ignored.

Accepted for publication September 8, 2004.

This study was supported by a grant from the Midwest Nursing Research Society (2003 Dissertation Grant).

The author acknowledges members of the dissertation committee: Shaké Ketefian, EdD, RN, FAAN, Chair and Professor, Director of the Office of International Affairs, and Violet Barkauskas, PhD, RN, FAAN, Associate Professor, both from the School of Nursing at the University of Michigan; Heather Laschinger, PhD, RN, Professor, Associate Director of Nursing Research and Chair of Graduate Nursing Programs, from the School of Nursing at the University of Western Ontario; and Lance Sandelands, PhD, Professor and Chair of the Department of Organizational Psychology, from the Department of Psychology at the University of Michigan. The guidance and support provided by all committee members are greatly appreciated.

Corresponding author: Milisa Manojlovich, PhD, RN, CCRN, School of Nursing, University of Michigan, 400 N Ingalls, Room 4306, Ann Arbor, MI 48103 (e-mail: mmanojo@umich.edu).
References


Fatigue and Fatigue-Relieving Strategies Used by Hong Kong Chinese Patients After Hemopoietic Stem Cell Transplantation

Winnie K. W. So ▼ Josepha W. M. Tai

Background: Fatigue, a common symptom for patients after hemopoietic stem cell transplantation (HSCT), can be relieved by specific strategies initiated by patients, provided the strategies are perceived as effective.

Objectives: To explore the intensity of fatigue experienced by patients after HSCT for the treatment of hematologic malignancy, and to determine the effectiveness of self-initiated fatigue-relieving strategies.

Methods: A self-administered survey was mailed to 284 Hong Kong Chinese patients who had been hospitalized for HSCT. The patients were required to provide their demographic data and complete the questionnaires of the Revised Piper Fatigue Scale—Chinese Version and the Fatigue Relief Scale—Chinese Version. The data obtained then were analyzed to assess the patients’ level of fatigue, the types of fatigue-relieving strategies used, and the effectiveness of the strategies.

Results: The participants (n = 157) returned the survey at a response rate of 70.9%. Most of the participants perceived a moderate level of fatigue (mean, 4.7 ± 1.7). More than 15% reported that they experienced a high level of fatigue. Most of the participants (95%) chose to take some rest and reduce physical activities. The five most effective strategies were sleeping, napping, lying down, being massaged, and stopping current activity.

Discussion: Most of the participants perceived the level of fatigue after HSCT as moderate, although various self-initiated fatigue-relieving strategies were used.

Key Words: fatigue • fatigue-relieving strategies • hemopoietic stem cell transplant

Fatigue is widely acknowledged as a common and distressing symptom of cancer and its treatment, yet fatigue experienced by patients undergoing hemopoietic stem cell transplantation (HSCT) is relatively undocumented (Cella, Lai, Chang, Peterman, & Slavin, 2002).

Patients managed by HSCT are conditioned with a high dose of chemotherapy or total body irradiation before they receive stem cells to suppress the immune system, minimize graft rejection, and reduce the recurrence of malignancy (Deeg, Klingemann, Phillips, & van Zant, 1999). Patients may attempt to alleviate fatigue symptoms with self-care measures (Graydon, Bubela, Irvine, & Vincent, 1995; Ream & Richardson, 1999; Richardson & Ream, 1997). Therefore, the efficacy of fatigue-relieving measures is vital to patients for fatigue relief and maintenance of function.

The findings of a survey on the intensity of fatigue among Hong Kong Chinese HSCT patients are reported. Whereas fatigue-relieving strategies have been studied in non-Chinese populations (Ream & Richardson, 1999), research on the efficacy of such strategies in Chinese populations has been limited. This study was conducted to explore the intensity of fatigue and the strategies used for its relief. The following specific research questions were investigated:

1. What is the intensity of fatigue experienced by the Hong Kong Chinese patients who have undergone HSCT?
2. Which fatigue-relieving strategies do these patients use?
3. How effective are the selected fatigue-relieving strategies?

Literature Review

Hematologic malignancy is one of the most common types of cancer among Hong Kong Chinese adults (Hong Kong Cancer Registry Highlight, 2001). Chemotherapy and radiotherapy destroy not only malignant cells, but also
normal cells, affecting the patients’ bone marrow. Complications of infection and hemorrhage also arise. Myelosuppression commonly limits the dose of traditional chemotherapy. Hemopoietic stem cell transplantation is used to “rescue” the marrow when the patients are conditioned with a high dose of chemotherapy or radiotherapy.

Two types of HSCT are used: (a) allogeneic HSCT, using stem cells (i.e., the marrow or peripheral blood) harvested from a human leukocyte antigen identical donor, and (b) autologous HSCT, using stem cells collected directly from the patient and returned later when needed (Deeg et al., 1999). However, after HSCT, the transplantation conditioning and immunosuppression cause significant morbidity. Fatigue was reported as a common symptom among long-term survivors (Tai & Lee, 2002).

Post-HSCT patients generally require nearly a year for full recovery of physical functioning. However, fatigue often extends beyond a year and influences patients as they return to accustomed activities (Andrykowski et al., 1997). Fatigue is associated with common illnesses, affecting daily life and the quality of life (Aaronson et al., 1999). Nevertheless, healthy people do not consider fatigue a serious problem because it can be relieved by a good night’s sleep. In contrast, chronic fatigue is unpredictable and unlikely to be relieved by rest (Johnston & Coward, 2001).

Fatigue is a complex and multidimensional concept associated with physical, emotional, and cognitive manifestations that affect a person’s well-being, daily activities, lifestyle, social and work-related activities, familial and sexual relations, and compliance with conventional therapy. Thus, fatigue has a direct impact on an individual’s quality of life (Cella, Peterman, Passik, Jacobsen, & Breitman, 1998). In addition, fatigue is described as a lack of energy, weakness, tiredness, exhaustion, lethargy, weariness, depression, inability to concentrate, malaise, sleepiness, lack of motivation, and decreased mental status (Winningsen et al., 1994). Piper (1993) defined fatigue as “an unusual, abnormal, or excessive whole body tiredness disproportionate to, or unrelated to, activity or exertion” (p. 279). Ream and Richardson (1996) defined fatigue as “a subjective and an unpleasant symptom which incorporates total body feelings ranging from tiredness to exhaustion creating an unrelenting overall condition which interferes with individuals’ ability to function to their normal capacity” (p. 527).

On the basis of the general definitions, a panel of experts in fatigue provided supplementary information and defined cancer-related fatigue as “an unusual, persistent, subjective sense of tiredness related to cancer or cancer treatment that interferes with usual functioning” (Mock, Piper, Sabbatini, & Escalante, 2000, p. 152). This definition helps to differentiate the unusual features of cancer-related fatigue from the fatigue experienced by healthy individuals, showing that cancer-related fatigue is more distressing and less likely to be relieved (Cella et al., 2002).

A set of diagnostic criteria for cancer-related fatigue proposed by Cella et al. (1998) includes the following: (a) at least 6 of 11 criteria such as significant fatigue, generalized weakness, diminished concentration, decreased motivation, short-term memory problems, and sleeping disturbance nearly every day during the same 2-week period in the preceding month; (b) symptoms of fatigue causing significant distress in social, occupational, and daily activities; (c) symptoms not primarily related to comorbidity with psychiatric disorders; and (d) evidence from physical examination and historical or laboratory findings that the symptoms are a consequence of cancer or its treatment.

Cancer-related fatigue has a profound negative impact on the individual’s quality of life overall, and special self-initiated strategies are used to relieve it. Various self-initiated strategies used by patients with cancer to relieve the intensity of fatigue have been reported in Western countries. A sample of 99 women undergoing either chemotherapy or radiation therapy for breast, cervical, endometrial, or ovarian cancer rested or reduced their level of activities (Graydon et al., 1995). Participants with ovarian, colorectal, or small cell lung cancer (n = 109) who were receiving chemotherapy adopted similar strategies (Richardson & Ream, 1997). It is further reported that 121 patients with breast, prostate, and ovarian cancer and lymphoma who were receiving radiation therapy frequently adopted other strategies such as sitting and sleeping (Irvine, Vincent, Graydon, & Bubela, 1998). These common sense strategies should relieve the associated fatigue.

Subsequent exercises have been evaluated and found to be effective in reducing the fatigue perceived by patients during treatment for cancer (Mock et al., 2001). Other alternative approaches such as education, attention–restoration, and psychosocial activities have been suggested for the relief of cancer-related fatigue (Ream & Richardson, 1999).

The Practice Guidelines for Fatigue, developed by a multidisciplinary panel of clinicians and researchers who represented 17 institutions of the National Comprehensive Cancer Network in 2000, were presented by Mock (2001). The guidelines proposed an algorithm in which patients were to be assessed regularly and treated according to their level of fatigue. If perceived fatigue was absent or present in a mild level of intensity, education and supportive care was provided. If fatigue was perceived to be moderate or severe, a more focused history and physical examination was conducted to evaluate associated factors. If no other factor was found, a more comprehensive assessment was made to determine the cause of fatigue.

Fatigue management usually started with basic education and counseling. Next came correction of causative factors, followed by the introduction of nonpharmacologic and pharmacologic approaches to fatigue reduction. Although the guidelines represented the current science of fatigue management, Mock (2001) noted that additional
research was needed to support the guidelines and fill up the knowledge gaps.

Recently the Beating Fatigue program, with elements of assessment and monitoring, education, coaching, and emotional support was developed to facilitate the management of fatigue among patients receiving chemotherapy (Ream, Richardson, & Alexandar-Dann, 2002). This program was evaluated by a pilot study with eight participants. The attitudes of the participants were positive, and they utilized the opportunity to share their beneficial strategies. The results of the study supported the view that a multifaceted approach is needed for the management of cancer-related fatigue (Ream et al., 2002).

Conceptual Framework
The aim of this study was to explore the level of fatigue and the effectiveness of self-initiated fatigue management for HSCT patients was guided by Orem’s Self-Care Deficit Theory (Foster & Bennett, 2002). Orem described self-care as an individual’s performance of activities to maintain health and well-being. Self-care was shown to involve requisites initiated by individuals that influence human functioning. According to Orem, self-care requisites comprise three groups characterized as (a) universal (associated with general well-being), (b) developmental (associated with human developmental processes), and (c) health devotional (associated with structural and functional deviations). When self-care is effectively performed, it maintains human functioning and contributes to human development (Foster & Bennett, 2002).

Orem’s Self-Care Deficit Theory structured this exploration and analysis of fatigue and fatigue-relieving strategies through the proposition that effective self-care helps to maintain function. Successful fatigue-relieving strategies are essential for symptomatic relief of fatigue. For example, maintenance of a balance between rest and activity is a universal requisite for reducing fatigue. When HSCT patients perceive fatigue, they search for self-initiated strategies that relieve fatigue. Orem’s self-care theory then can guide the study of such fatigue, especially the strategies associated with its relief.

Methods
Design and Sample
A descriptive, cross-sectional design was used. The study was conducted at a large teaching hospital in Hong Kong that offered HSCT and was the first in Hong Kong to use sibling and matched but unrelated donors successfully (Liang, 1999). As of May 2003, 912 HSCTs had been performed in this hospital.

All the patients who had undergone HSCT in this teaching hospital were recruited. A sample list of these patients was prepared at the transplantation center using a computerized program. To ensure privacy and confidentiality, only authorized personnel were allowed to access the information. The sample inclusion criteria required that the participants were 18 years or older when they underwent HSCT, had received a diagnosis of hematologic malignancy, had completed the course of HSCT and had been discharged home, were able to read and write Chinese, and could be reached by telephone.

Of the 284 potential participants identified from the list of patients, 220 met the inclusion criteria after the initial contact by telephone. The remainder included 26 who were unreachable and 38 who were either hospitalized or deceased. Of the eligible participants, 157 returned the questionnaires, giving a response rate of 70.9%.

Procedure
Approval for the study was obtained from the hospital institutional review board. Eligible participants received a mail package containing an invitation letter explaining the purposes of the study; an information sheet indicating the details of the study, the confidentiality of the participant’s response, and the need and intention of the researchers to contact the participant by telephone; an informed consent form for the participant to sign; an encoded questionnaire; and a stamped addressed envelope for the return of the signed consent form and questionnaire. The questionnaires were coded for subsequent follow-up discussion in case unclear responses had to be clarified. The researchers retained the participants’ contact information and destroyed it to ensure confidentiality once it was deemed unnecessary.

Instruments
The self-reported survey consisted of three parts: demographic data, the Revised Piper Fatigue Scale—Chinese version (RPFS-CV), and the Fatigue Relief Scale—Chinese version (FRS-CV).

Demographic Data
The demographic data were obtained by self-report, and additional data such as medical diagnosis, conditioning regimen, and length of the posttransplantation period were obtained from the participants’ medical records.

Revised Piper Fatigue Scale—Chinese Version
The RPFS, a multidimensional assessment tool, was effective in differentiating the fatigue perceived between two different groups and within the same group at different intervals (Monga, Kerrigan, Thornby, & Monga, 1999). The RPFS consists of 22 items that measure the overall level of fatigue perceived on a visual analog scale of values ranging from 0 to 10. It also measures four dimensions of fatigue: severity, affective, sensory, and cognition. Scores of 0 to 3
represent no fatigue or a mild level of fatigue; scores greater than 3 and up to 6 represent a moderate level of fatigue; and scores greater than 6 represent a high level of fatigue (Mock, 2001). In this study, the no fatigue or mild fatigue group was further divided into a no fatigue group with scores of zero and a mild fatigue group with the scores greater than 0 but less than 3. The Cronbach’s alpha coefficient for the level of fatigue perceived by survivors of breast cancer was greater than .92 for each subscale and .91 for the entire scale (Piper et al., 1998).

Because the participants in this study were Hong Kong Chinese patients, the Revised Piper Fatigue Scale—Chinese Version (RPFS-CV) was used, with Cronbach’s alpha coefficients ranging from .89 to .93 for each subscale and .91 for the entire scale (So, Dodgson, & Tai, 2003). The RPFS-CV was used to minimize the language barriers in assessment (Varricchio, 1997).

Fatigue Relief Scale—Chinese Version The Fatigue Relief Scale (FRS) was first developed by Graydon et al. (1995) to assess the strategies used by patients with cancer to manage the level of fatigue they perceived and the efficacy of the strategies. The FRS consists of 15 strategies grouped into four dimensions: (a) reducing or ceasing activities (e.g., lie down, nap, sit, sleep, and stop activity), (b) increasing physical or social activities (e.g., exercise, socialize, and walk), (c) using distraction (e.g., listen to music, read, and watch TV), and (d) others (e.g., ask for help, do something different, eat or drink, and work).

The participants were asked to indicate which of these strategies they had used to relieve the intensity of fatigue. They also were asked to indicate on a 100-mm linear analog scale with a range of 0 to 100 the effectiveness of the strategies. A mean efficacy score of the strategies used by the participants was calculated to measure the level of fatigue relieved. This score was obtained by calculating the effectiveness ratings of the strategies used by the participants and dividing by the total number of strategies used. The higher the score, the greater the patient’s relief from the intensity of fatigue. The test–retest reliability according to Graydon et al. (1995) was .7, with reference to 51 participants over a period of 24 hours.

Because the participants were Hong Kong Chinese, the Fatigue Relief Scale-Chinese Version (FRS-CV) was developed, and a pilot study was conducted initially to determine whether the tool was readable and understandable by the Chinese participants. The proposed items were reviewed by 17 patients within the target population, and comments were provided on their meaning and clarity.

A panel of experts also reviewed the translated items to ensure the content validity (Carlson, 2000). The following modifications of the FRS-CV were made. “Massage” was added to the “others” dimension because it is one of the methods used in traditional Chinese medicine to stimulate blood circulation (Chan & Molassiotis, 2000). The techniques of massage used by this study group also were included, and these often were performed by other people using simple maneuvers such as rubbing, kneading, and pressing (Wright, 1995). “Work,” “ask for help,” and “read” were removed because “work” could be resumed only about a year after transplantation, and the patients were reluctant to “ask for help” because self-reliance is characteristic of the Chinese culture (Bond, 1991). Yet, not all participants had received formal education. A visual analog scale with scores ranging from 0 (not effective) to 10 (very effective) was used instead of a 100-mm linear analog scale because the 100-mm scale was too difficult for the participants to comprehend. Retest was performed after 2 weeks, resulting in a test–retest reliability of .83. The Cronbach’s alpha coefficient for each dimension ranged from .64 to .82, and the alpha for entire scale was .80.

Statistical Analysis Descriptive analysis was performed on all the variables. The participants were grouped according to their perceived level of fatigue: Group 1 (no fatigue perceived), Group 2 (mild level of perceived fatigue), Group 3 (moderate level of perceived fatigue), and Group 4 (high level of perceived fatigue). One-way analysis of variance (ANOVA) was used to determine whether there were any significant differences in the level of fatigue perceived among the four groups. The Pearson correlation coefficient was used to examine the relation between the mean fatigue scores of the participants and the degree of fatigue relieved. Those who did not experience fatigue were not required to analyze the fatigue-relieving strategies frequently used. A p value of .05 or less was regarded as statistically significant (Munro, 2001).

Results Demographics and Intensity of Fatigue The demographic data and original diagnoses of the participants are tabulated in Table 1. The total mean RPFS-CV scores and the subscale scores are tabulated in Table 2. The overall score for the intensity of fatigue was compared with different demographic data. Significant differences were found in the categories of age (p < .01), marital status (p < .05), employment status (p < .01), and household income (p < .01). The participants more likely to perceive a high level of fatigue were older, married, not employed, and poorer. Details of the findings and discussions are reported elsewhere (So et al., 2003).

The participants were grouped according to four levels of fatigue. Group 1 (n = 43, 27.4%) did not perceive any fatigue and Group 4 (n = 25, 15.9%) perceived a high level of fatigue. An ANOVA procedure then was used to determine the significant differences in the levels of fatigue perceived among the four groups (p < .01). The results of the post hoc Games–Howell test showed that the mean scores for each subscale and the mean total score for intensity of fatigue in Group 4 were significantly higher than in other three groups (p < .01).

Furthermore, the length of the posttransplantation period was comparatively shorter for the patients who experienced a higher level of fatigue (M = 43.5 months; r = 2–107 months) than for those who experienced no fatigue or a mild level of fatigue (M = 53.7 months; r = 4–136 months) or a moderate level of fatigue (M = 53.3 months; r = 2–133 months), although the results were not
Fatigue and Fatigue-Relieving Strategies

Nursing Research  January/February 2005   Vol 54, No 1

statistically significant (So et al., 2003). These findings showed that patients perceived a lower level of fatigue as time increased than when they had received their treatments (Irvine et al., 1998).

Fatigue-Relieving Strategies Frequently Used and Their Effectiveness

The participants who did not experience fatigue (n = 43) were excluded from the analysis of the types of strategies frequently used and the effectiveness of these strategies. Thus, the data collected from 105 participants were used for further analysis.

The five most frequently used fatigue-relieving methods were related to rest (e.g., lie down, sit, sleep, and nap) and a reduction in activities (e.g., stop what you are doing), whereas the five most effective methods were similar to the five most frequently used except for “massage” and “sit down.” In other words, “sit down” was frequently used to relieve the intensity of fatigue, but it was not effective, whereas “massage” was rated as effective, but was not frequently used.

Nevertheless, there was no relation between the overall score for the intensity of fatigue and the mean efficacy of fatigue-relieving strategies. The strategies are tabulated in Table 3, and the most effective strategies are shown in Table 4.

Discussion

Intensity of Fatigue

Although fatigue is a common symptom for patients after HSCT, the respondents in this survey experienced only moderate levels of fatigue. Perhaps the length of the post-transplantation period and the retrospective design influenced their experienced level of fatigue. A smaller subgroup with distinct demographic characteristics including

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>M (SD)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>39.9 (9.9)</td>
<td>18–69</td>
</tr>
<tr>
<td>Months after BMT</td>
<td>52.9 (36.7)</td>
<td>2–136</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>84</td>
<td>53.5</td>
</tr>
<tr>
<td>Female</td>
<td>73</td>
<td>46.5</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>95</td>
<td>60.5</td>
</tr>
<tr>
<td>Single</td>
<td>57</td>
<td>36.3</td>
</tr>
<tr>
<td>Separated / divorced</td>
<td>4</td>
<td>2.5</td>
</tr>
<tr>
<td>Widow / widower</td>
<td>1</td>
<td>.6</td>
</tr>
<tr>
<td>Education level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No formal education</td>
<td>2</td>
<td>1.3</td>
</tr>
<tr>
<td>Primary</td>
<td>27</td>
<td>17.2</td>
</tr>
<tr>
<td>Secondary</td>
<td>79</td>
<td>50.3</td>
</tr>
<tr>
<td>Tertiary</td>
<td>49</td>
<td>31.2</td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not employed</td>
<td>85</td>
<td>54.1</td>
</tr>
<tr>
<td>(unemployed, housewife, retired)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed (full time and part time)</td>
<td>72</td>
<td>45.9</td>
</tr>
<tr>
<td>Income per month (Hong Kong $)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;= 15,000 (ie, &lt;= US $1,923)</td>
<td>66</td>
<td>44.3</td>
</tr>
<tr>
<td>15,001–30,000 (ie, US $1,923–3,846)</td>
<td>48</td>
<td>32.2</td>
</tr>
<tr>
<td>&gt; 30,000 (ie, &gt; US $3,846)</td>
<td>35</td>
<td>23.5</td>
</tr>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronic myelocytic leukemia</td>
<td>50</td>
<td>35.2</td>
</tr>
<tr>
<td>Acute myeloid leukemia</td>
<td>27</td>
<td>19.0</td>
</tr>
<tr>
<td>Acute lymphocytic leukemia</td>
<td>12</td>
<td>8.5</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>10</td>
<td>7.0</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>9</td>
<td>6.3</td>
</tr>
<tr>
<td>Myelodysplastic syndrome</td>
<td>9</td>
<td>6.3</td>
</tr>
<tr>
<td>Hodgkin’s disease</td>
<td>7</td>
<td>4.9</td>
</tr>
<tr>
<td>Others</td>
<td>12</td>
<td>8.4</td>
</tr>
<tr>
<td>Type of treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allogeneic</td>
<td>112</td>
<td>78.3</td>
</tr>
<tr>
<td>Autologous</td>
<td>28</td>
<td>19.6</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>2.1</td>
</tr>
<tr>
<td>Conditioning regime</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>99</td>
<td>73.3</td>
</tr>
<tr>
<td>Bu-Cy (120)</td>
<td>50</td>
<td>37.0</td>
</tr>
<tr>
<td>CBV</td>
<td>12</td>
<td>8.9</td>
</tr>
<tr>
<td>Bu-Cy (150)</td>
<td>8</td>
<td>5.9</td>
</tr>
<tr>
<td>Carbo PEC</td>
<td>7</td>
<td>5.2</td>
</tr>
<tr>
<td>BEAM</td>
<td>4</td>
<td>3.0</td>
</tr>
<tr>
<td>Others</td>
<td>18</td>
<td>13.2</td>
</tr>
<tr>
<td>Chemo-radiotherapy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cy-TBI</td>
<td>21</td>
<td>15.6</td>
</tr>
<tr>
<td>Bu-Cy-TBI</td>
<td>10</td>
<td>7.4</td>
</tr>
<tr>
<td>Fludarabine-TBI</td>
<td>3</td>
<td>2.2</td>
</tr>
<tr>
<td>Others</td>
<td>2</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Note. BMT = bone marrow transplant; Bu-Cy (120) = busulfan + cyclophosphamide 120 mg/kg; Bu-Cy (150) = busulfan + cyclophosphamide 150 mg/kg; CBV = cyclophosphamide, BCNU / carmustine and VP 16; Carbo PEC = carboplatin + etoposide + cyclophosphamide; BEAM = BCNU/carmustine + etoposide + cytarabine + melphalan; Bu-Cy-TBI = busulfan + cyclophosphamide and total body irradiation; Cy-TBI = cyclophosphamide and total body irradiation; Fludarabine-TBI = fludarabine and total body irradiation.
those significantly older, who often were not employed or had a lower income, experienced higher levels of fatigue at the time of data collection (So et al., 2003). Because the study used a retrospective cross-sectional design, the period between the patients’ treatment and data collection varied among the participants. This feature likely affected the level of fatigue reported and was, in turn, affected by the patients’ recovery processes. The sample was not homogeneous because the patients were assessed in different months or years after transplantation.

The high internal consistency of the entire PFS-CV scale reflected the higher level of fatigue the patients perceived and the greater difficulty they encountered when they engaged in daily activities. Also, distinct characteristics (i.e., older, married, not employed, having a lower income) associated with the high level of fatigue supported Orem’s propositions that conditioning factors do affect the ability of individuals to perform self-care.

### Fatigue-Relieving Strategies Frequently Used and Their Effectiveness

The five fatigue-relieving strategies frequently used were related to rest and reduction in activities. These strategies were natural responses to fatigue and indicated that the patients had self-care ability to relieve the symptoms when fatigue became disruptive to their daily lives (Richardson

### Table 2. Revised Piper Fatigue Scale—Chinese Version (RPFS-CV) Scores Grouped by Level of Fatigue

<table>
<thead>
<tr>
<th></th>
<th>All Participants</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Number of participants (%)</td>
<td>157 (100)</td>
<td>43 (27.4)</td>
<td>17 (10.8)</td>
<td>63 (40.1)</td>
<td>25 (15.9)</td>
</tr>
<tr>
<td>Sensory</td>
<td>4.6 (2.3)</td>
<td>.0a</td>
<td>2.4b (1.0)</td>
<td>5.0c (1.2)</td>
<td>7.6d (1.4)</td>
</tr>
<tr>
<td>Mood</td>
<td>4.2 (2.0)</td>
<td>.0a</td>
<td>2.4b (1.3)</td>
<td>4.5c (1.3)</td>
<td>6.6d (1.3)</td>
</tr>
<tr>
<td>Behavior</td>
<td>4.5 (2.3)</td>
<td>.0a</td>
<td>2.2b (1.3)</td>
<td>4.0c (1.7)</td>
<td>7.0d (1.4)</td>
</tr>
<tr>
<td>Affect</td>
<td>4.6 (1.9)</td>
<td>.0a</td>
<td>2.9b (1.3)</td>
<td>4.3c (1.5)</td>
<td>6.6d (1.6)</td>
</tr>
<tr>
<td>Total RPFS-CV score</td>
<td>4.7 (1.7)</td>
<td>.0a</td>
<td>2.5b (1.6)</td>
<td>4.4c (1.9)</td>
<td>6.9d (1.9)</td>
</tr>
</tbody>
</table>

Note. Group 1 = no fatigue perceived group (total fatigue scores = 0); Group 2 = mild level of fatigue perceived group (total fatigue scores ≥ 0–3); Group 3 = moderate level of fatigue perceived group (total fatigue scores ≥ 3–6); Group 4 = high level of fatigue perceived group (total fatigue scores ≥ 6); Mean with different superscripts (a, b, c, d) differs significantly at p < .01 in the post hoc Games-Howell test. Missing data = 9.

### Table 3. Frequently Used Strategies for the Relief of Fatigue (N = 105)

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Five Most Common Methods Perceived by the Participants</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lie down</td>
<td>100 (95.2)</td>
</tr>
<tr>
<td>2</td>
<td>Sit down</td>
<td>98 (93.3)</td>
</tr>
<tr>
<td>3</td>
<td>Nap</td>
<td>97 (92.4)</td>
</tr>
<tr>
<td>4</td>
<td>Sleep</td>
<td>96 (91.4)</td>
</tr>
<tr>
<td>5</td>
<td>Stop current activity</td>
<td>96 (91.4)</td>
</tr>
<tr>
<td>6</td>
<td>Watch TV</td>
<td>94 (89.5)</td>
</tr>
<tr>
<td>7</td>
<td>Eat and drink</td>
<td>92 (87.6)</td>
</tr>
<tr>
<td>8</td>
<td>Listen music</td>
<td>89 (84.8)</td>
</tr>
<tr>
<td>9</td>
<td>Walk</td>
<td>80 (76.2)</td>
</tr>
<tr>
<td>10</td>
<td>Massage</td>
<td>79 (75.2)</td>
</tr>
<tr>
<td>11</td>
<td>Do other things</td>
<td>78 (74.3)</td>
</tr>
<tr>
<td>12</td>
<td>Exercise</td>
<td>76 (72.4)</td>
</tr>
<tr>
<td>13</td>
<td>Social activities</td>
<td>70 (66.7)</td>
</tr>
</tbody>
</table>

### Table 4. Effective Strategies for the Relief of Fatigue (N = 105)

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Five Most Effective Methods Perceived by the Participants</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sleep</td>
<td>8.1 (2.2)</td>
</tr>
<tr>
<td>2</td>
<td>Nap</td>
<td>7.9 (2.0)</td>
</tr>
<tr>
<td>3</td>
<td>Lie down</td>
<td>7.7 (1.9)</td>
</tr>
<tr>
<td>4</td>
<td>Massage</td>
<td>7.0 (2.1)</td>
</tr>
<tr>
<td>5</td>
<td>Stop what you’re doing</td>
<td>6.7 (2.3)</td>
</tr>
<tr>
<td>6</td>
<td>Sit down</td>
<td>6.6 (2.0)</td>
</tr>
<tr>
<td>7</td>
<td>Listen to music</td>
<td>6.1 (2.3)</td>
</tr>
<tr>
<td>8</td>
<td>Walk</td>
<td>5.5 (2.4)</td>
</tr>
<tr>
<td>9</td>
<td>Eat and drink</td>
<td>5.4 (2.3)</td>
</tr>
<tr>
<td>10</td>
<td>Exercise</td>
<td>5.3 (2.7)</td>
</tr>
<tr>
<td>11</td>
<td>Watch TV</td>
<td>5.2 (2.2)</td>
</tr>
<tr>
<td>12</td>
<td>Do other things</td>
<td>5.0 (2.3)</td>
</tr>
<tr>
<td>13</td>
<td>Social activities</td>
<td>4.8 (2.2)</td>
</tr>
</tbody>
</table>
& Ream, 1997). These strategies may be effective for acute fatigue (Aaronson et al., 1999), but for cancer-related fatigue, these strategies reduce functional capacity and decrease exercise tolerance (Nail & Winningham, 1995). Ultimately, the intensity of fatigue may be reduced further when these strategies are used (Wells & Fedric, 2001). Healthcare professionals often recommend these fatigue-relieving strategies (Mock et al., 1997), indicating lacking knowledge of the pathophysiology of fatigue.

In this study, the participants frequently chose to rest and reduce their activities to relieve the intensity of fatigue. Resting and napping were continuously reported for some samples in North America (Graydon et al., 1995; Mock, 2001). A detailed cultural comparison of fatigue management for HSCT recipients in Western and Asian populations was not found. Cultural beliefs and patterns likely influence the preferences for fatigue-relieving strategies and the manner in which they are used in self-care.

The participants in this study understood that after transplantation, their health condition was comparatively poor and included low immunity (So et al., 2003). They preferred staying at home to reduce their activities, to restore their health, and to prevent the “bad winds” from entering the body. In Chinese culture, “wind illness” is a traditional belief that illness stems from an imbalance in the forces of yin and yang. This imbalance often is caused by eating certain kinds of food such as “cold food” (e.g., melons and cold dishes), or by exposure to the external environment (e.g., draft and “cold air”) (Leininger, 1995). Because Hong Kong is polluted and overcrowded, patients are reluctant to go outside and participate in outdoor activities.

Some fatigue-relieving strategies used by the respondents were perceived as effective. The “sit down” strategy was used frequently, but not effectively, whereas “massage” was regarded as effective, but not frequently used. Massage has been used by various populations to reduce muscle tension, improve blood circulation, and promote relaxation (Ahles et al., 1999). Various studies have showed that massage therapy relieves fatigue and associated psychological symptoms of distress (Ahles et al., 1999). The respondents in the current study did not frequently use massage therapy to relieve the intensity of fatigue although the strategy was rated effective.

The disjuncture between the effectiveness of massage in fatigue management and its infrequent use in this sample raises questions about the role of massage and cultural values of self-care for fatigue among Chinese HSCT patients. On the other hand, most Chinese practice self-restraint to enhance living in harmony (Gao, Ting-Toomey, & Gudykunst, 1996). Under the influence of Confucianism and Buddhism, the Chinese developed an indigenous concept of retribution, “bao,” which requires a person not to owe other persons favors and directs each person to make an effort toward repaying what favors they do owe. This tradition renders the choice of using massage therapy unpopular because the person does not know whether the favors of others can be repaid, and this would make it difficult to sustain a harmonious personal relationship (Leung, 1996).

The frequently used strategies in this study were a variety of self-care strategies reflecting the choice of the participants despite the intensity of the fatigue they perceived or how long ago they had undergone HSCT. The different kinds of self-care strategies may be related to the Confucian tradition of self-discipline. Evaluation of the effectiveness of fatigue-relieving strategies culturally should address their influence on social harmony and the psychological state of the individuals (Phillips & Pearson, 1996).

This study involved several limitations to the internal and external validity that shapes the interpretation of the findings. The design of the study was retrospective and cross-sectional, using a convenience sample. Additionally, the pattern, consistency, and intensity of fatigue experienced by the patients over time were not evaluated. The period between HSCT and data collection varied considerably and may have influenced the perceived level of fatigue.

Fatigue and the effectiveness of fatigue-relieving strategies among the patients who had undergone HSCT for hematologic malignancy were examined in this study. Generally, the patients perceived a moderate level of fatigue after transplantation. They usually chose the self-care measures of rest and reduction in physical activities to relieve the intensity of fatigue. The fatigue-relieving strategies most frequently used by the patients were not necessarily the most effective ones. “Sit down” was one of the strategies frequently used by the participants, but it was not very effective, whereas “massage” was rated effective, but not frequently used. Future studies should explore the effectiveness of fatigue-relieving strategies such as balancing rest and exercise, involving the family members, and implementing massage therapy. ▼

Accepted for publication June 29, 2004.

This study was funded by the Research Council of the University of Hong Kong.

The authors thank Drs. Sarah Kagan and Judith Payne for reviewing the manuscript; Dr. Joan Dodgson and Ms. May Ho for data analysis; Miss Carmen Yam, Miss Debbie Wong, and Mr. Vicky Chan for their contributions in data collection and data entry work; Miss Agnes Ng and Miss S. M. Tang for their assistance in literature review; and the help of the nursing staff of the bone marrow transplantation unit during the process of data collection.

Corresponding author: Winnie K.W. So, MHA, BN, RN, Department of Nursing Studies, University of Hong Kong, 4/F, Academic and Administration Block, Faculty of Medicine Building, 21 Sassoon Road, Pokfulam, Hong Kong SAR, China (e-mail: wksusu@hkusu.hku.hk).

References


Combining Evidence in Nursing Research

Methods and Implications

Robin Whittemore

Methods

Combining Evidence in Nursing Research

Methods and Implications

Robin Whittemore

Published reports include the following titles: “A synthesis of the evidence,” “A review of the literature,” “A systematic review,” “A meta-analysis,” or “An integrative review.” Are all of these research reviews similar or are there differences in the methods, purpose, and thus clinical applicability? The purpose of this presentation is to identify the defining characteristics of different research review methods, the common stages of all research reviews, and criteria for evaluating quality. Greater clarity regarding research review methods has the potential to enhance methodological rigor and subsequently the applicability of findings to practice or policy.

Defining Characteristics of Research Review Methods

Methods to conduct research reviews in the health professions have been utilized since the 1970s to synthesize findings from numerous primary studies and allow for increased generalizability about a phenomenon of interest (Jackson, 1980). Research reviews summarize past research critically to draw conclusions, informing research, practice, and policy.

Conducting a quality research review, however, is quite challenging. Traditional research reviews have been criticized for a lack of precision and high potential for subjectivity and bias (Beck, 1999; Cooper, 1998). The multiplicity of research methods of individual studies, the quantity of studies on any potential topic, and the complexity of findings complicate the task of completing a quality research review (Beck, 1999; Brown, Upchurch, & Acton, 2003). Statistical methods have been developed, yet

Robin Whittemore, PhD, APRN, is Associate Research Scientist and Lecturer, Yale School of Nursing, New Haven, Connecticut.
cannot be utilized with all types of reviews. Methods continue to evolve so as to improve the rigor of the process; however, rapid development of methods over the last 10 years has contributed to the current methodological confusion.

Methods for research reviews in nursing include integrative, meta-analytic, systematic, and qualitative reviews. Although there are common features to these methods, each method has a distinct purpose, sampling frame, definition, and type of analysis (Table 1).

**Integrative Reviews**

Integrative reviews are the broadest category of research reviews and can encompass empirical or theoretical literature, or both, depending on the purpose of the research. Integrative reviews can be focused on methodology, theory, or the results of differing empirical studies with a wide range of implications, Table 2 (Broome, 1993). One of the distinct advantages to the integrative review approach is the ability to combine data from different types of research designs and include theoretical as well as empirical literature. Although the inclusion of multiple research designs can complicate the analysis, greater variety in the sampling frame has the potential to increase the depth and breadth of conclusions. The richness of the sampling frame also can contribute to a comprehensive portrayal of the topic of interest.

**Meta-Analysis**

Meta-analysis is an approach to combining the evidence of multiple primary studies that employs statistical methods, thus increasing the objectivity of findings (Glass, 1976). The sampling frame is narrow because the research design and hypotheses of primary studies have to be very similar, if not identical (Cooper, 1998). With this approach, each primary study is abstracted, coded, and assembled into a quantitative database. Findings are transformed into a common metric to calculate an overall effect and the magnitude of the effect. Adjustment for sample size and study quality can be incorporated also into the analysis, which is a significant advantage to this method (Broome, 1993; Oxman & Guyatt, 1988).

In the absence of a large randomized clinical trial, meta-analyses have been identified as a good source of evidence for evidence-based practice and, therefore, are being undertaken more frequently (Acton, 2001; Friedland, Shlipak, Subak, Bent, & Mendelson, 1998). However, the requirement for highly related or identical hypotheses in primary studies precludes the use of this type of method for all questions aimed at combining the evidence of multiple studies.

**Systematic Review**

Systematic reviews are an approach to combining the evidence of multiple studies that differ from the

---

### TABLE 1. Methods of Nursing Research Reviews

<table>
<thead>
<tr>
<th>Type of Review and Exemplar</th>
<th>Definition</th>
<th>Purpose</th>
<th>Scope</th>
<th>Sampling Frame</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrative review (Redeker, 2000)</td>
<td>A summary of the literature on a specific concept or content area whereby the research is summarized, analyzed, and overall conclusions are drawn</td>
<td>To review methods, theories, and/or empirical studies around a particular topic</td>
<td>Narrow or broad</td>
<td>Quantitative or qualitative research; theoretical literature; methodological literature</td>
<td>Narrative</td>
</tr>
<tr>
<td>Meta-analysis (Clemmens, 2001)</td>
<td>A summary of past research using statistical techniques to transform findings of studies with related or identical hypotheses into a common metric and calculating the overall effect, the magnitude of effect, and subsample effects</td>
<td>To estimate the effect of interventions or relationships</td>
<td>Narrow</td>
<td>Quantitative research of similar methodology</td>
<td>Statistical</td>
</tr>
<tr>
<td>Systematic review (Forbes, 1998)</td>
<td>A summary of past research using an objective and rigorous approach of studies with related or identical hypotheses</td>
<td>To summarize evidence regarding a specific clinical problem</td>
<td>Narrow</td>
<td>Quantitative research of similar methodology</td>
<td>Narrative or statistical</td>
</tr>
<tr>
<td>Metasummary Metasynthesis Formal grounded theory Metastudy (Beck, 2002)</td>
<td>A summary of past research combining the findings from multiple qualitative studies</td>
<td>To inform research or practice by summarizing processes or experiences</td>
<td>Narrow or broad</td>
<td>Qualitative research</td>
<td>Narrative</td>
</tr>
</tbody>
</table>
TABLE 2. Types of Integrative Reviews

<table>
<thead>
<tr>
<th>Type of Review</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methodological</td>
<td>Critical review and analysis of the designs, methods, and analyses in a series of studies (i.e., methodological problems in the research of children with a chronic illness)</td>
</tr>
<tr>
<td>Theoretical</td>
<td>Critical review of theories around a particular topic examining supportive and nonsupportive evidence</td>
</tr>
<tr>
<td>Empirical</td>
<td>Critical review of empirical studies (quantitative or qualitative, or both) around a particular topic examining outcomes and relationships between variables</td>
</tr>
</tbody>
</table>

Note. Adapted from Broome (1993).

aforementioned approaches in their purpose. Systematic reviews inform practice by summarizing evidence regarding a specific clinical problem and are the focus of evidence-based practice initiatives (Cochrane Collaboration, n.d.). A well-specified clinical question, use of the best available evidence, explicit methods, and an exhaustive search for relevant primary studies are hallmarks of this method (Counsell, 1997; Forbes, 2003; Greenhalgh, 1997). Systematic reviews may include the statistical methods of meta-analysis if primary studies meet the assumptions required for meta-analyses. If primary studies cannot be combined statistically, a narrative analysis similar to the integrative review method is undertaken. Thus far, systematic reviews have combined the evidence of quantitative primary studies with similar designs; however, methods for including qualitative research are currently being developed (Cochrane Qualitative Research Methods Group, 2002).

Qualitative Methods

Methods to combine qualitative research have proliferated in the last decade (Jensen & Allen, 1996; Kearney, 2001; Paterson, Thorne, Canam, & Jillings, 2001; Sandelowski & Barroso, 2003; Sandelowski, Docherty, & Emden, 1997). These distinct approaches use qualitative research as the sampling frame; however, they differ in analysis and levels of interpretation. Metasummaries are descriptive and provide summaries of the findings from multiple primary qualitative studies (Sandelowski and Barroso), Metasynthesis, metastudies, formal grounded theory, and metaethnography methods propose to critically analyze primary studies and synthesize findings into a new theory or overarching framework on the topic of interest. Although qualitative research is contextual and complex to combine, combining the evidence from multiple studies has the potential to broaden the generalizability and clinical applicability of small sample qualitative designs.

Stages of Research Reviews

Although differences with respect to definition, purpose, scope, sampling frame, and analysis of review methods have been proposed, all research reviews are considered research of research and, therefore, should meet the same standards as primary research in methodological rigor (Cooper, 1998; Ganoung, 1987). Each method includes a problem formulation stage, a literature search stage, a data evaluation stage, a data analysis stage, and a presentation stage (Cooper).

Problem Identification Stage

The initial phase of any research endeavor is a clear identification of the problem and the associated review purpose. Subsequently, the method and the variables of interest (i.e., target population, health care problem or intervention, clinical outcomes) are determined. The selection of the review method depends on the clinical problem and the state of the science regarding that topic. For example, a clinical problem about conflicting treatment options where numerous randomized clinical trials have been undertaken would best be addressed by a systematic review or a meta-analysis. A clinical problem about conflicting evidence of family functioning in chronic illness whereby only descriptive research (quantitative and qualitative) has been utilized would best be addressed by an integrative review.

Literature Search Stage

The literature search stage is a critical element to conducting a quality research review because incomplete and biased searches result in an inadequate database for the review and the potential for faulty conclusions (Conn, Isaramalai, et al., 2003; Cooper, 1998). Ideally, any review of past research includes all of the relevant literature on the specified problem; yet, obtaining this literature is challenging and potentially costly (Jadad, Moher, & Klassen, 1998). Computerized databases are cost-effective, especially if assisted by a librarian; however, limitations associated with inconsistent search terminology and indexing problems may yield only part of the eligible studies (Conn, Isaramalai, et al., 2003). In addition, computerized databases search for published literature, which more often includes statistically significant findings or what is referred to as publication bias (Soeken & Sriputusanap, 2003). Publication bias may overestimate any treatment effect or relationship between variables by as much as 30% (McAuley, Pham, Tuggwell, & Moher, 2000). In contrast, unpublished studies identified through conference abstract proceedings or through networking may not have the same methodological rigor as peer-reviewed publications (Conn, Valentine, Cooper, & Rantz, 2003). Including unpublished studies in research reviews remains controversial and at this time they are often not included. Ancestry searching, journal hand-searching, networking, and searching research registries are approaches also recommended for searching the literature (Conn, Isaramalai, et al., 2003).
A comprehensive search for a research review identifies the maximum number of eligible studies, utilizing as many search strategies as resources allow (at least 2–3 strategies; Jadad et al., 1998). In the research review report, the literature search process should clearly document the search terms, the databases used, the search strategies, and the inclusion criteria for determining relevant studies. Inclusion criteria generally identify the time delimits of the review, the sampling frame (type of research designs), the targeted population, the treatment or intervention (if applicable), and the outcomes of interest (conceptual and operational definitions).

**Data Evaluation Stage**

Once the relevant literature has been collected, common data are extracted from primary studies for subsequent analysis. Reliable and valid coding procedures are essential to ensure methodological rigor (Broome, 1993; Brown et al., 2003). In addition to coding data pertinent to the research review, methodological features of primary studies are also extracted to evaluate the overall quality of individual studies. Quality is determined by evaluating the internal validity of primary studies or “the extent to which study design, conduct, and analysis systematically avoid or minimize potential sources of bias” (Conn & Rantz, 2003, p. 323). Evaluating the quality of differing designs is complex and, therefore, no gold standard exists. Quality assessment ideally examines generic methodological criteria (Table 3) as well as criteria specific to a particular field (i.e., control for particular confounding variables) (Greenhalgh, 1997; West et al., 2002). To ensure accuracy, often two independent reviewers code individual studies for content and quality, and interrater agreement is calculated.

Once quality scores have been calculated for the primary studies of a research review, how these scores are incorporated into the data analysis is somewhat controversial (Conn & Rantz, 2003; Jadad et al., 1998). The relationship between study quality and study outcomes has been conflicting. For example, some investigators have reported that low quality scores underestimate effect sizes whereas others have reported that low quality scores overestimate effect sizes (Balk et al., 2002; Moher, Pham, et al., 1998). Despite this uncertainty, quality evaluation of primary studies is highly recommended and can be incorporated into the research review in several ways (Conn & Rantz, 2003). Quality evaluation scores can be used as an inclusion criteria for selecting relevant primary studies for the review. Quality scores can also be displayed graphically according to specific outcomes of studies and

### TABLE 3. Study Constructs Believed to Affect the Quality of Primary Studies

<table>
<thead>
<tr>
<th>Study Construct</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>Inclusion and exclusion criteria well defined and appropriate</td>
</tr>
<tr>
<td></td>
<td>Health, demographic, socioeconomic status, and other confounding</td>
</tr>
<tr>
<td></td>
<td>characteristics considered</td>
</tr>
<tr>
<td>Study protocol</td>
<td>Study procedures systematic and well specified</td>
</tr>
<tr>
<td>Measurement</td>
<td>Choice of outcomes appropriate</td>
</tr>
<tr>
<td></td>
<td>Data collection specified and systematic</td>
</tr>
<tr>
<td></td>
<td>Measures demonstrate adequate reliability and validity</td>
</tr>
<tr>
<td>Attrition</td>
<td>Withdrawals, dropouts, or other losses from study identified and</td>
</tr>
<tr>
<td></td>
<td>accounted for</td>
</tr>
<tr>
<td>Threats to validity</td>
<td>Confounders and bias carefully considered and controlled</td>
</tr>
<tr>
<td>Statistical analysis</td>
<td>Statistics appropriate and well described</td>
</tr>
<tr>
<td></td>
<td>Levels of significance or confidence intervals, or both, reported</td>
</tr>
<tr>
<td></td>
<td>Intention-to-treat analysis for longitudinal data</td>
</tr>
<tr>
<td>Discussion</td>
<td>Conclusions supported by results with possible biases and limitations</td>
</tr>
<tr>
<td>Intervention (if applicable)</td>
<td>Detailed information about treatment, setting, and interventionist</td>
</tr>
<tr>
<td></td>
<td>provided</td>
</tr>
<tr>
<td></td>
<td>Efforts undertaken to ensure treatment integrity</td>
</tr>
<tr>
<td></td>
<td>Blinding (of patients, investigators, data collectors, and care</td>
</tr>
<tr>
<td></td>
<td>providers)</td>
</tr>
<tr>
<td></td>
<td>Randomization of subjects to treatment and control group</td>
</tr>
<tr>
<td></td>
<td>Evidence of study group comparability at baseline provided</td>
</tr>
</tbody>
</table>

Note. Adapted from Lohr & Carey (1999); West et al. (2002).
considered when analyzing variability of results (i.e., outliers). For meta-analysis reviews, quality scores are often incorporated into the analysis by weighing studies on the basis of quality or by comparing the results of studies by differing quality scores (Jadad et al., 1998).

**Data Analysis Stage**

Data analysis in research reviews requires that the investigator order, categorize, and summarize data from primary individual studies into a unified conclusion about the research problem (Cooper, 1998). A thorough and unbiased interpretation of individual studies with clear, useful conclusions are the goals of this stage (Slavin, 1995). Data analysis techniques and procedures will vary depending on the purpose and the particular type of research review.

With the exception of meta-analysis, all research reviews employ a narrative or qualitative analysis. Coded data of each individual study are compared with all other studies for similarities and differences around variables of interest. Visual aids such as tables or schematic representations of the relationships among codes, concepts, and ideas often are employed (Miles & Huberman, 1994). Outliers are identified and examined for reasons for variability explored (i.e., population, setting, outcomes, design; Oxman & Guyatt, 1988). The potential for subjectivity and bias in the data analysis phase of research reviews requires careful attention. Explicit records of all coding structures and data analysis decisions must be maintained to increase the rigor and the transparency of the process. Conclusions must have clear supportive evidence from the findings of all individual studies (Moher, Jadad, & Klassen, 1998).

Statistical procedures have been developed to enhance the objectivity of the data analysis stage of research reviews (Table 4). However, there are limitations with respect to the review purpose and sampling frame for which these procedures are applicable. As previously mentioned, primary studies must have identical research designs, address a highly related, if not identical, research hypothesis with similar independent and dependent variables, and report specific statistics that can be extracted for data analysis (Cooper, 1998). The advantage to the incorporation of statistics in research reviews is the ability to estimate the magnitude of the effect in addition to the ability to examine the influence of other variables on the outcomes of interest quantitatively (i.e., quality of individual studies, gender). Statistical power and objectivity are increased with the use of statistical methods; however, accuracy of results remains dependent on diligence with the aforementioned stages of the process (Beck, 1999; Dunkin, 1996).

**Presentation of Results Stage**

Results of research reviews can be presented as a summary, an analysis, or a synthesis. Summaries are similar to descriptive research because they describe and display findings of primary studies by differing categories or themes. An analysis provides some summary and descriptive information; however, these reviews also provide critique of methods, outcomes, and/or applicability to practice. Results reported as a synthesis are the highest level of abstraction and include the creation of a new model or organizing framework for the problem or topic of interest.

Results of primary studies used in the review should be reported in detail so that readers may assess the basis for the conclusions that are drawn (Oxman & Guyatt, 1988; Slavin, 1995). Generally, tables document the evidence of individual studies, providing key details without superfluous data (Stetler et al., 1998). Major design characteristics, sample characteristics, variables, and outcomes are often displayed (i.e., sample size, type of design, setting). The provision of explicit data from primary studies allows the reader to ascertain that the conclusions of the review did not

---

**TABLE 4. Common Statistics and Graphic Displays in Meta-Analysis**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Definition</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect size</td>
<td>( d ) index reported when means of two groups are compared</td>
<td>Small effect if ( d = .20 ) or ( r = .10 )</td>
</tr>
<tr>
<td></td>
<td>( r ) index reported when describing the relationship between two variables</td>
<td>Medium effect if ( d = .50 ) or ( r = .30 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Large effect if ( d = .80 ) or ( r = .50 )</td>
</tr>
<tr>
<td>Risk</td>
<td>Odds ratio reported when both variables are dichotomous</td>
<td>Odds ratio of 1 means that the null hypothesis is true or the odds are the same in both conditions</td>
</tr>
<tr>
<td>Vote count</td>
<td>Estimation of statistical significance or direction, or both, of findings</td>
<td>Somewhat imprecise</td>
</tr>
<tr>
<td>Number to treat (NTT)</td>
<td>Number of patients needed to treat with the experimental treatment to prevent one patient from adverse outcome</td>
<td>Provide some indication of cost/benefit ratio</td>
</tr>
<tr>
<td>Forest plots</td>
<td>Visual plot of trial point estimates and confidence interval of individual study. Last line of plot represents synthesis of all studies</td>
<td>Vertical line represents no effect. Left side of vertical line indicates benefit of experimental treatment. Right side of vertical line indicates benefit of control condition</td>
</tr>
</tbody>
</table>

Note. Adapted from Cooper (1998).
exceed the evidence and the appropriate reasons for variability have been explored (Oxman, 1994).

**Quality of Research Reviews**

The process of combining the evidence of multiple primary studies is complex. Bias and error can occur at any stage of the process as each method involves numerous decisions and judgments (Dunkin, 1996; Oxman, 1994). For example, the literature search may not be comprehensive if valid studies are omitted. Data from primary studies can be incorrectly extracted or interpreted and the data analysis may not accurately reflect all of the evidence. Thus, attention to quality is required at all stages of a research review (Table 5).

Quality in research reviews begins with a clearly defined problem, purpose, and method. Because of the complexity associated with completing a review without bias, reviews ideally are conducted by a team with expertise in the content area and the methodology (Gregson, Meal, & Avis, 2002). A written protocol is developed prior to the beginning of the review and specifies the research review plan in detail. The search protocol and the eligibility criteria for primary studies are clearly stated and the results of the search are comprehensive (using 2–3 search strategies). The quality of primary studies should be evaluated and incorporated in the analysis and interpretation of findings (Conn & Rantz, 2003). Studies ideally are assessed for their ability to be combined logically or statistically (heterogeneity tests), or both ways. The data extraction process should be explicit, unbiased, and reproducible (Friedland et al., 1998). A table of key elements of each individual study ideally is included and enough detail is provided to evaluate that conclusions are based on the evidence. If possible, graphic displays of data are also included (i.e., plotting quality of studies with a particular outcome). The analysis stage should be explicit, with conclusions demonstrating thoughtful consideration of all studies. In addition, reasons for variability are explored. Results ideally capture the complexity of the clinical problem and are interpreted within the context of current health care. Last, methodological limitations of the research review are articulated.

The ongoing process of knowledge development in a practice discipline requires the integration of primary research into a more comprehensive understanding of the topic of interest. Methods to combine the evidence of multiple primary studies have advanced considerably over the last decade. Although the systematic review is the approach most commonly associated with evidence-based practice initiatives, all review methods have the potential to have direct applicability to nursing practice. Quasi-experimental, descriptive, and qualitative research is important to nursing science and practice. Greater clarity regarding research review methods as well as articulating quality criteria for all methods has the potential to enhance methodological rigor, confidence in conclusions, and subsequently the applicability of findings from all types of research reviews to nursing practice.

---

**TABLE 5. Quality Criteria in Research Reviews**

<table>
<thead>
<tr>
<th>Quality Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Well-defined problem and review purpose</td>
</tr>
<tr>
<td>2. Explicit identification of review method</td>
</tr>
<tr>
<td>3. Investigators with expertise in content and methodology</td>
</tr>
<tr>
<td>4. Clear specification of review process and protocol</td>
</tr>
<tr>
<td>5. Comprehensive and explicit literature search</td>
</tr>
<tr>
<td>6. Explicit, unbiased, and reproducible data extraction for content and quality</td>
</tr>
<tr>
<td>7. Primary study quality considered in analysis</td>
</tr>
<tr>
<td>8. Data analysis is systematic and variability of findings addressed</td>
</tr>
<tr>
<td>9. Evidence included from primary studies</td>
</tr>
<tr>
<td>10. Conclusions based on evidence and capture complexity of clinical problem</td>
</tr>
<tr>
<td>11. Methodological limitations identified</td>
</tr>
</tbody>
</table>

---

**References**


Conn, V. S., Isaramalai, S., Rath, S., Jan-
62
tarakupt, P., Wadhawan, R., & Dash, Y. (2003). Beyond MEDLINE for litera-
ture searches. *Journal of Nursing Scholar-
ship*, 35, 177-182.
Conn, V. S., Valentine, J. C., Cooper, H. M.,
& Rantz, M. J. (2003). Grey literature in meta-
Cooper, H. (1998). *Synthesizing research: A
Counsell, C. (1997). Formulating ques-
tions and locating primary studies for in-
Dunkin, M. J. (1996). Types of errors in synthe-
aging the behavioural symptomatology asso-
Forbes, D. A. (2003). An example of the use of systematic reviews to answer an effec-
Friedland, D. J., Shlipak, M. G., Subak, L.
L., Bent, S. W., & Mendelson, T. (1998). Evaluating integrative litera-
ture. In D.J. Friedland, A.S. Go, & J. B.
Davoren (Eds.), *Evidence-based medicine: A framework for clinical practice* (pp. 221-246). Stamford, CT: Appletorn & Lange.
Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *The Educationa-
cational Researcher*, 5, 1083-1088.
Greenhalgh, T. (1997). Papers that sum-
marize other papers (systematic reviews
Gregson, P. R., Meal, A. G., & Avis, M.
Jackson, G. B. (1980). *Methods for inte-
preting systematic reviews: II. How did the authors find the studies and assess their quality? *Archives of Pediatric & Adolescent Medicine*, 152, 812-817.
synthesis of qualitative findings. *Qualitative Health Research*, 6, 553-
560.
153.
Lohr, K. N., & Carey, T. S. (1999). Assess-
ing “best evidence”: Issues in grading the quali-
Mcauley, L., Pham, B., Tugwell, P., & Moher,
D. (2000). Does the inclusion of grey litera-
ture influence estimates of intervention effectiveness reported in meta-
Moher, D., Jadad, A. R., & Klassen, T. P.
(1998). Guides for reading and interpret-
ing systematic reviews: III. How did the authors synthesize the data and make their conclusions? *Archives of Pediatrics and Adolescent Medicine*, 152, 915-920.
Moher, D., Pham, B., Jones, A., Cook, D.
(1998). Does quality of reports of ran-
domized trials affect estimates of inter-
vention efficacy reported in meta-
233.
Sandelowski, M., Docherty, S., & Emden, C. (1997). Focus on qualitative meth-
Slovin, R. E. (1995). Best evidence synthesis: An intelligent alternative to meta-
Soeken, K. L., & Sripusanapan, A. (2003). Assessing publication bias in meta-
analysis. *Nursing Research*, 52, 57-60.
206.
West, S., King, V., Carey, T. S., Lohr, K. N.,
McKay, N., Sutton, S. F., et al. (2002). *Systems to rate the strength of scientific evidence* (AHRQ Publication No. 02-